

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

ABRAHAM COHEN
THE JOHNS HOPKINS UNIVERSITY

F. D. MURNAGHAN
THE JOHNS HOPKINS UNIVERSITY

T. H. HILDEBRANDT
UNIVERSITY OF MICHIGAN

J. F. RITT
COLUMBIA UNIVERSITY

R. L. WILDER
UNIVERSITY OF MICHIGAN

WITH THE COÖPERATION OF

OYSTEIN ORE
H. P. ROBERTSON
M. H. STONE
T. Y. THOMAS
G. T. WHYBURN

E. T. BELL
H. B. CURRY
E. J. MCSHANE
HANS RADEMACHER
OSCAR ZARISKI

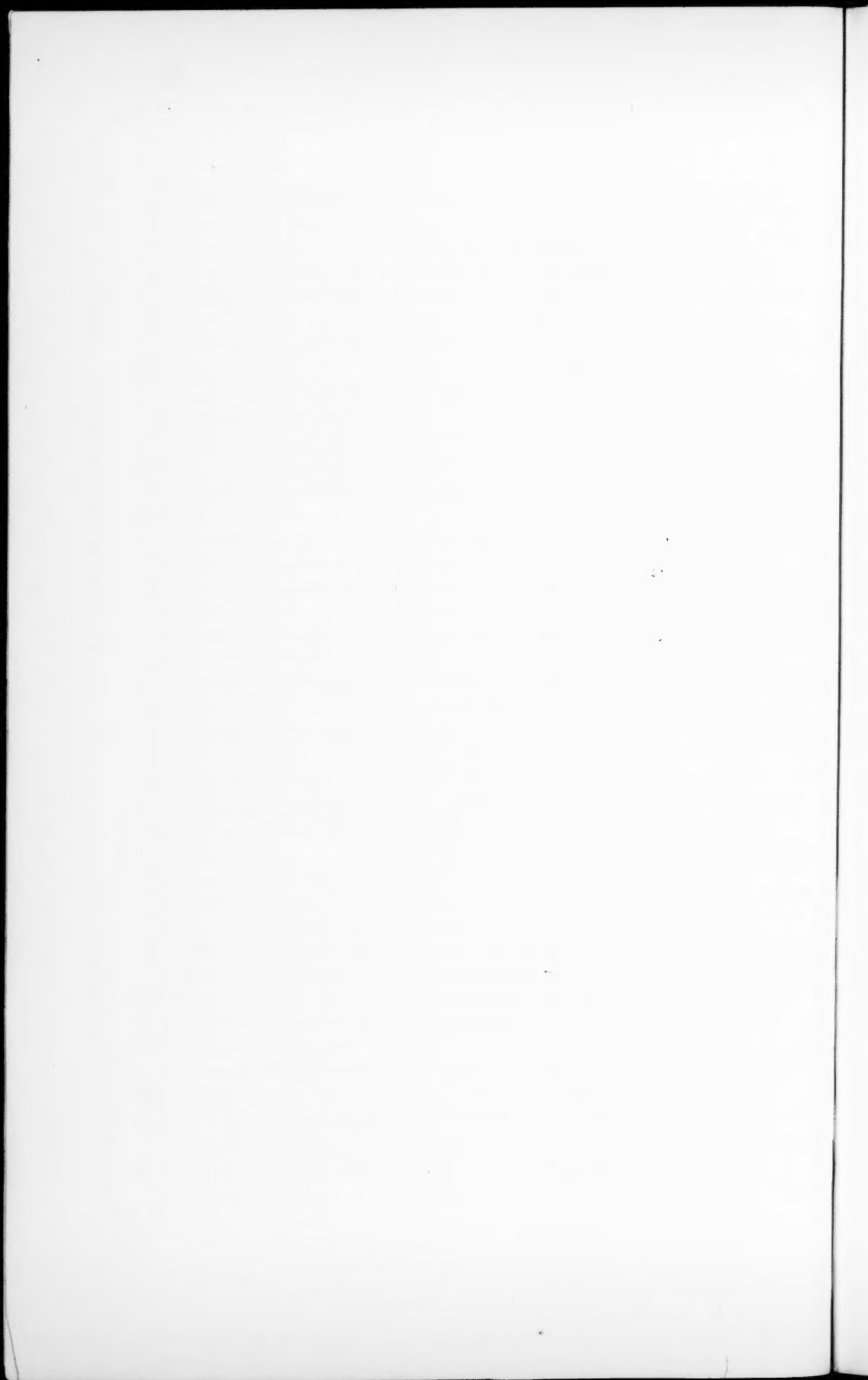
C. R. ADAMS
R. D. JAMES
SAUNDERS MACLANE
GABOR SZEGÖ
LEO ZIPPIN

PUBLISHED UNDER THE JOINT AUSPICES OF
THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

VOLUME LXII

1940

THE JOHNS HOPKINS PRESS
BALTIMORE, MARYLAND
U. S. A.



SYMBOLIC DYNAMICS II. STURMIAN TRAJECTORIES.*

By MARSTON MORSE and GUSTAV A. HEDLUND.

1. Introduction. In a recent paper¹ we initiated a theory of symbolic dynamics. In this theory we consider unending sequences of symbols or *symbolic trajectories* and devote attention to those properties of symbolic trajectories which are suggested by dynamical considerations. A symbolic trajectory is formed from symbols taken from a finite set of generating symbols subject to certain rules of admissibility. In SD admissibility conditions were formulated of such generality that the resulting symbolic trajectories include in particular those which arise in the geodesic problem on surfaces which satisfy the condition of uniform geodesic instability.

However, no surface of the topological type of a torus satisfies the condition of uniform geodesic instability and the admissibility conditions of SD do not include those which arise in the case of the torus.

In the present paper we consider a class of symbolic trajectories formed from two generating symbols subject to admissibility conditions defined by a simple comparison property. These are the symbolic trajectories which characterize the geodesics on a flat torus. They may be used to characterize the distribution of the zeros of the solutions of a differential equation of the form $y'' + f(x)y = 0$, where $f(x)$ is a periodic function of x . We term the trajectories of this class *Sturmian*. A first fundamental result is as follows:

Sturmian trajectories possess certain numerical characteristics, namely, a frequency, a pole, and a type index, and admit mechanical constructions uniquely determined by these characteristics.

There are three types of Sturmian trajectories,—irrational, skew and periodic. The trajectories of irrational type are recurrent but not periodic; those of skew type are not recurrent. The recurrency function of a recurrent Sturmian trajectory is completely determined by the frequency α of the trajectory and may be denoted by $R(n, \alpha)$. We introduce the variable $\gamma = \alpha(1 + \alpha)^{-1}$. Let C_ν/D_ν be the convergents in a continued fraction representation of γ . We have the following fundamental theorem:

*Received June 19, 1939.

¹ Cf. Morse and Hedlund. (References will be found in the bibliography at the end of the paper.) This paper will hereafter be referred to as SD. Numerous references will be found in the bibliography at the end of SD.

When α is irrational, $R(n, \alpha)$ increases by unity when n increases from $n-1$ to n except when $n = D_v$, $v = 0, 1, \dots$. For these exceptional values of n we have the relation

$$R(D_v, \alpha) = D_{v+1} + 2D_v - 1 \quad (v = 0, 1, 2, \dots)$$

starting with $v = 1$ in the special case $D_0 = D_1$.

The preceding theorem thus gives a simple mode of evaluating $R(n, \alpha)$ when α is irrational. Previous to this the only non-periodic recurrent trajectory of which the recurrency function had been determined was the Morse recurrent trajectory (cf. SD § 8).

The evaluation of $R(n, \alpha)$ permits various extensions of our knowledge of recurrency functions. In particular we are able to solve one of the problems posed at the end of SD. We had shown in SD § 7 that if $R(n)$ is the recurrency function of a general non-periodic recurrent trajectory, $\liminf_{n \rightarrow \infty} R(n)/n \geq 2$. The results of the present paper show that the constant 2 cannot be replaced by a greater constant.

The proper choice of α yields a recurrency function $R(n, \alpha)$ such that

$$R(n, \alpha) < \frac{5 + \sqrt{5}}{2} n$$

with $R(n, \alpha)$ becoming infinite more slowly than for any other previously known non-periodic trajectory. A final result on the asymptotic behavior of $R(n, \alpha)$ is as follows. Let $\phi(x)$ be a positive monotonically increasing function of x defined for $x > 0$. As n becomes infinite the lim. sup. of

$$\frac{R(n, \alpha)}{n\phi(\log n)}$$

is finite or infinite for almost all values of α according as the series

$$\sum_{n=1}^{\infty} \frac{1}{\phi(n)}$$

converges or diverges. In particular the lim. sup. of

$$\frac{R(n, \alpha)}{n \log n}$$

is infinite for almost all values of n while the lim. sup. of

$$\frac{R(n, \alpha)}{n (\log n)^c} \quad (c > 1)$$

is zero for almost all values of α .

The results of this paper and its predecessor will presently be given its appropriate dynamical setting (cf. G. D. Birkhoff) in terms of trajectories on a space form.

I. CLASSIFICATION AND REPRESENTATION.

2. The comparison condition and general theorems. We shall consider sequences X of two symbols a and b of the forms

$$(2.1) \quad \cdots aB_{-1}aB_0aB_1a \cdots,$$

$$(2.2)' \quad \cdots aB_r aB_{r+1}a \cdots,$$

$$(2.2)'' \quad \cdots aB_{r-1}aB_r a,$$

$$(2.3) \quad aB_r a \cdots aB_s a \quad (r \leq s),$$

in which B_n is a finite block of b 's. We admit that B_n may be the null set. We term B_n the *cell* of X of index n , and term X a *cell-sequence*. A cell-sequence X of the form (2.1), (2.2) or (2.3) will be respectively termed a *cell-series*, a *cell-beam* or a *chain*. A chain which contains n cells will be termed an *n-chain*. If the chain (2.3) appears in X it will be termed the chain $[r, s]$ of X .

Two cell-series (cell-beams) X and Y will be regarded as *identical* if and only if the cells of X and Y have the same index range and if cells with the same index are identical. On the other hand, two n -chains of the form

$$\begin{aligned} aB_{p+1}a \cdots aB_{p+n}a, \\ aB'_{q+1}a \cdots aB'_{q+n}a, \end{aligned}$$

will be regarded as *identical* if and only if

$$B_{p+i} = B'_{q+i} \quad (i = 1, 2, \cdots, n).$$

The number of symbols b in an n -chain x will be called the *b-length* of x . We shall be concerned with cell-sequences X which satisfy the following condition.

C. Under Condition C the *b-lengths* of any two n -chains of X with the same n shall differ by at most one.

We term this condition the *comparison condition*. Cell-sequences which satisfy the comparison condition will be called *Sturmian*. As we shall see, Sturmian sequences appear in the theory of linear second order differential equations.

THEOREM 2.1. The *b-lengths* b_m and b_n of arbitrary m - and n -chains of a Sturmian chain satisfy the relation

$$(2.4) \quad n(b_m + 1) > m(b_n - 1).$$

We shall give an inductive proof of the theorem, first noting that (2.4) holds when $m = n = 1$. Let N be any positive integer at most the maximum of m and n in the theorem. We assume the truth of (2.4) for n and m both less than N , and shall prove that (2.4) holds when m and n are at most N . Since (2.4) holds for $m = n$, there remain two cases to consider.

Case I. $N = m > n$. We here set $m = pn + q$ with $0 \leq q < n$. An m -chain x may be regarded as a sequence of p successive n -chains followed by a q -chain, two successive chains having a symbol a in common. The b -lengths of our p n -chains are each at least $b_n - 1$ so that

$$(2.5) \quad b_m \geq p(b_n - 1) + b_q$$

where b_q is the b -length of the final q -chain of x . We are assuming that (2.4) holds when m and n are both less than N , so that

$$(2.6) \quad n(b_q + 1) > q(b_n - 1).$$

Adding 1 to both members of (2.5) and multiplying by n we find that

$$(2.7) \quad n(b_m + 1) \geq np(b_n - 1) + n(b_q + 1).$$

Upon using (2.6), (2.7) takes the form (2.4) and the proof is complete in Case I.

Case II. $N = n > m$. Set $n = pm + q$ with $0 \leq q < m$. Essentially as in Case I we find that

$$(2.8) \quad b_n \leq p(b_m + 1) + b_q,$$

where b_n is the b -length of an arbitrary n -chain y and b_q is the b -length of the final q -chain of y . By virtue of our inductive hypothesis

$$(2.9) \quad m(b_q - 1) < q(b_m + 1).$$

Subtracting 1 from both members of (2.8), then multiplying by m and using (2.9), relation (2.4) results again as in Case I.

The proof of the theorem is complete.

THEOREM 2.2. *If b_n is the b -length of an arbitrary n -chain of a Sturmian beam or series, then b_n/n tends to a finite limit α as n becomes infinite.*

It follows from (2.4) that

$$\left| \frac{b_n}{n} - \frac{b_m}{m} \right| < \frac{1}{m} + \frac{1}{n},$$

and the theorem follows directly.

We term α the *frequency* of the Sturmian beam or series. When $\alpha \neq 0$ we set $\beta = 1/\alpha$ and term β the corresponding *rotation number*. When $\alpha = 0$, β shall be ∞ by convention.

When X is a Sturmian chain we shall set

$$(2.10) \quad \alpha' = \max \left[\frac{b_m}{m} - \frac{1}{m} \right], \quad \alpha'' = \min \left[\frac{b_m}{m} + \frac{1}{m} \right],$$

b_m ranging over all b -lengths of m -chains of X .

THEOREM 2.3. (a) *A necessary and sufficient condition that a cell-sequence T be Sturmian is that there exist a constant $\alpha \geq 0$ such that the b -lengths b_n of n -chains of T satisfy one of the two following sets of conditions for each n :*

$$(2.11)' \quad n\alpha - 1 < b_n \leq n\alpha + 1,$$

$$(2.11)'' \quad n\alpha - 1 \leq b_n < n\alpha + 1.$$

(b) *If T is a Sturmian chain, conditions (2.11) are satisfied if and only if α is on the interval $\alpha' \leq \alpha \leq \alpha''$, the right and left equalities prevailing in (2.11) at most when $\alpha = \alpha''$ and α' respectively.* (c) *If T is a Sturmian beam or series, (2.11) is satisfied if and only if α is the frequency.*

The conditions (2.11)' or the conditions (2.11)'' are sufficient that T be Sturmian since there are at most two integral values of b_n which satisfy (2.11)' or (2.11)'' respectively for a given n , and these integral values differ by at most one.

To prove the conditions (2.11) necessary we suppose T Sturmian.

We begin with the case in which T is a chain. It follows from (2.4) that

$$(2.12) \quad \frac{b_m}{m} - \frac{1}{m} < \frac{b_n}{n} + \frac{1}{n},$$

so that $\alpha' < \alpha''$. Moreover $\alpha' \geq 0$ except in the trivial case in which $b_m = 0$ for each m . It is easily seen that (2.11) holds for $\alpha' < \alpha < \alpha''$. For in such a case

$$(2.13) \quad n\alpha + 1 > n\alpha' + 1 \geq n \left[\frac{b_n}{n} - \frac{1}{n} \right] + 1 = b_n.$$

Similarly

$$(2.14) \quad n\alpha - 1 < n\alpha'' - 1 \leq n \left[\frac{b_n}{n} + \frac{1}{n} \right] - 1 = b_n.$$

For $\alpha' < \alpha < \alpha''$, (2.11) thus holds with the equalities excluded. It is also clear from (2.13) and (2.14) that (2.11)' holds when $\alpha = \alpha'$ and (2.11)'' when $\alpha = \alpha''$ but that (2.11)' does not hold in general when $\alpha = \alpha''$ nor (2.11)'' when $\alpha = \alpha'$.

The preceding analysis includes a proof of (b), as well as a proof of the necessity of (2.11) when T is a finite chain.

We come to the case in which T is a Sturmian beam or series with frequency α . That $b_m \leq m\alpha + 1$ follows at once from (2.12) upon letting n become infinite. Similarly we see from (2.12) that $b_n \geq n\alpha - 1$ upon letting m become infinite. The conditions (2.11) taken as a whole are accordingly satisfied by T . But it is impossible that $b_m = m\alpha + 1$ and $b_n = n\alpha - 1$ for the same beam or cell-series T . For in such a case we find that

$$m(b_n + 1) = n(b_m - 1),$$

contrary to (2.4). Thus conditions (2.11) are necessary in one of the two forms.

That (2.11) holds at most when α is the frequency follows upon dividing the respective members of (2.11) by n and letting n become infinite.

The proof of the theorem is complete.

3. The classification of Sturmian beams and series according to frequencies. Let R be a Sturmian beam with a rational frequency α . When $\alpha > 0$ we set $\alpha = q/p$ where q and p are relatively prime integers. When $\alpha = 0$ we understand that $q = 0$ and $p = 1$.

LEMMA 3.1. *In a Sturmian beam R with a rational frequency q/p , there cannot exist two p -chains with b -lengths different from q , with different cell indices.*

The lemma is illustrated by the Sturmian series

$$(3.1) \quad \cdots aB_{-1}aB_0aB_1a \cdots$$

in which the cell-series obtained by omitting B_0a is periodic with cell lengths alternately 2 and 3, starting with $b(B_1) = 3$. The b -length of B_0 shall be 3. Here $p = 2$, $q = 5$. The 2-chains in general have the b -length $q = 5$. But the 2-chain aB_0aB_1a has the b -length 6.

We come to the proof of the lemma.

The b -lengths b_p of p -chains of R satisfy (2.11)' or (2.11)". We consider the case in which (2.11)' is satisfied. Then b_p must be q or $q + 1$.

We suppose the lemma false. There then exist two p -chains x and y of R with b -lengths $q + 1$ and with different cell indices. Without loss of generality we can suppose that x precedes y in R . Let w be the m -chain of R with x as its initial p -chain and y its terminal p -chain. We distinguish two cases: Case I. $m \geq 2p$; Case II. $p < m < 2p$.

Case I. Let z be the subchain of w whose cells are not cells of x or y , and suppose that z is an r -chain. We understand that r may be zero. Let b_r and b_m be the b -lengths of z and w respectively. We have $m = 2p + r$ and

$b_m = b_r + 2q + 2$. Upon applying (2.11)' to b_r and to b_m respectively, we find that

$$(3.2) \quad \frac{rq}{p} - 1 < b_r \leq \frac{rq}{p} + 1,$$

$$(3.3) \quad (2p + r) \frac{q}{p} - 1 < b_r + 2q + 2 \leq (2p + r) \frac{q}{p} + 1.$$

Upon subtracting $2q$ from each member of (3.3) we see that (3.2) is satisfied with b_r replaced by $b_r + 2$. Since this is impossible we infer that Case I is impossible.

Case II. Let z be the subchain of w whose cells occur in both x and y , and suppose that z is an r -chain. Let b_r and b_m be the b -lengths of z and w respectively. We have $m = 2p - r$ and $b_m = 2q + 2 - b_r$. Upon applying (2.11)' to b_r and b_m respectively, we obtain (3.2) and the relation

$$(3.3)' \quad (2p - r) \frac{q}{p} - 1 < 2q + 2 - b_r \leq (2p - r) \frac{q}{p} + 1.$$

Upon formally adding the respective members of (3.2) and (3.3)' we obtain the relation

$$-2 < 2 \leq 2,$$

from which we infer that the equality holds in (3.2). Hence r must be a multiple of p . But this is impossible if $p < r < 2p$. Thus Case II is equally impossible.

The case where (2.11)'' holds is similarly treated, and the proof is complete.

A Sturmian beam or series will be said to have the *cell-period* p if its cells satisfy the relation

$$(3.4) \quad B_{i+p} = B_i$$

for each admissible i .

THEOREM 3.1. *A periodic Sturmian series T or beam R with rational frequency $\alpha = q/p$, where $^2 (q, p) = 1$, has the minimum cell-period p . The b -lengths b_n of its n -chains satisfy the condition*

$$(3.5) \quad n\alpha - 1 < b_n < n\alpha + 1,$$

assuming each integral value b_n which satisfies (3.5).

The p -chains of T have the constant b -length q . Otherwise there would be infinitely many p -chains with different cell-indices and with b -lengths different from q contrary to Lemma 3.1. Hence T has the cell-period p .

² The notation $(q, p) = 1$ shall mean that q and p are relatively prime integers.

Let s be an arbitrary cell-period of T and let r be the b -length of an s -chain of T . Then

$$\alpha = \frac{r}{s} = \frac{q}{p}.$$

This is possible only if s is a multiple of p . Hence p is the minimum cell-period of T .

That (3.5) is satisfied will follow from (2.11) once the equality signs are excluded from (2.11). But an equality can hold in (2.11) only if $n\alpha$ is an integer or zero, and this implies that n is a multiple of p . When $n = rp$, $b_n = rq$ since a p -chain of T has the b -length q , and we conclude that $b_n = n\alpha$. Hence the equality never prevails in (2.11) and (3.5) holds as stated.

To see that b_n assumes each integral value which satisfies (3.5) we first note that when n is a multiple of p , $n\alpha$ is the only value of b_n which satisfies (3.5). There remains the case where n is not a multiple of p . But if for the given n , b_n had but one value, n would be a period of T , and hence a multiple of p , contrary to hypothesis. Hence b_n assumes each integral value which satisfies (3.5).

A similar proof applies to beams.

The proof of the theorem is complete.

Sturmian series with irrational frequencies will be termed *irrational*.

THEOREM 3.2. *The b -lengths b_n of the n -chain of an irrational Sturmian series with frequency α satisfy the condition*

$$(3.6) \quad n\alpha - 1 < b_n < n\alpha + 1,$$

assuming each integral value b_n which satisfies (3.6).

The numbers b_n satisfy (2.11)' or (2.11)'' as we have seen. But when α is irrational the equality can never prevail in (2.11). Moreover, for each n , b_n assumes the two values defined by (3.6). Otherwise T would have the cell-period n , and hence a rational frequency.

The proof of the theorem is complete.

Sturmian series which have rational frequencies but which are not periodic will be termed *skew*. The appropriateness of the term will appear later. An example of a skew Sturmian series has already been given in (3.1). Another example T^* will be given here. To define T^* we first define a cell-series Z . The beam of Z whose first cell is B_m shall have the cell-period 5. The b -lengths of its cells shall have a period block 21211. The beam of Z whose final cell is B_{m-1} shall also have the cell-period 5. The b -lengths of its cells shall have a period block 11212. The b -lengths of cells of Z thus form a sequence

... (11212)(11212)(21211)(21211) ...

To obtain T^* from Z we replace B_m in Z by a cell of b -length 1. The series T^* is seen to be Sturmian. Its frequency is $7/5$.

To analyze skew Sturmian series we introduce several terms. An n -chain of a Sturmian series T whose b -length is the maximum or minimum among b -lengths of n -chains of T will be said to be of *max* or *min* type respectively. A cell B_r of T will be said to be of *max* or *min* type if the chain $aB_r a$ is of *max* or *min* type respectively.

It follows from Lemma 3.1 that in any skew Sturmian series T with frequency q/p there exists one and only one p -chain whose b -length is different from q . This chain will be called the *critical chain* of T .

THEOREM 3.3. *Let T be a skew Sturmian series with critical p -chain C . The beam following (preceding) the initial (final) cell of C has the cell-period p . The initial cell B of C is identical with the final cell of C while the cells immediately preceding and following C are identical and opposite in type to B .*

Let X and Y be respectively the beams preceding the final and following the initial cell of C . The beams X and Y have the cell-period p since each of their p -chains has the b -length q .

Suppose for simplicity that C is of *max* type. Let B' be the terminal cell of C . The beams

$$(3.7) \quad aBX, \quad YB'a$$

are not periodic for their terminal p -chains C have b -length $q+1$. All the p -chains in the two beams (3.7) will have the b -length q provided their terminal cells are reduced by a unit in b -length, following which reduction both beams have the cell-period p . The cells thereby replacing B and B' have copies in T and must be of minimum type. Hence B and B' are of maximum type and identical.

That the cell preceding (following) C in T is of type opposite to B follows from the fact that the beams (3.7) are not periodic.

The case where C is of minimum type is similarly treated.

THEOREM 3.4. *The b -length b_n of the n -chains of a skew Sturmian series T with frequency α satisfy one of the conditions (2.11). Condition (2.11)' {(2.11)''} is satisfied if the critical chain is of *max* type {*min* type}. The integers b_n assume all integral values satisfying (2.11)' {(2.11)''}.*

That the numbers b_n satisfy one of the conditions (2.11)' or (2.11)'' follows from Theorem 2.3. Let us suppose that the conditions (2.11)' are satisfied. Since T is not periodic, b_n must assume two distinct values for each

positive integer n . Corresponding to a given integer n there are only two integral values which satisfy (2.11)' and it follows that b_n must assume all integral values which satisfy (2.11)'. If $\alpha = q/p$, where $(q, p) = 1$, then $b_p = q$ or $q + 1$. The critical p -chain of T must have b -length $q + 1$ and hence is of max type.

Analogous arguments hold if the integers b_n satisfy (2.11)'' and the proof of the theorem is complete.

CONDITION A. *A set of m -chains will be said to satisfy Condition A if, n being any positive integer not exceeding m , the b -lengths of the sub n -chains of the given set of m -chains assume at most two values.*

LEMMA 3.2. *If a set of m -chains satisfies Condition A, the number of chains in the set cannot exceed $m + 1$.*

If X and Y are m - and n -chains respectively, XY shall mean the $(m + n)$ -chain whose first m -chain and last n -chain are X and Y respectively.

The lemma is obviously true if $m = 1$. We assume the lemma true for integers not exceeding $m - 1$ and prove that the lemma holds for the integer $m > 1$. If the lemma were not true there would exist a set of $m + 2$ different m -chains satisfying Condition A. Let us denote such a set by

$$C_1, C_2, \dots, C_{m+2}.$$

By the hypothesis of the induction, this set contains at most two different 1-chains B and B^* and therefore the set can be written in the form

$$(C) \quad C_1 = D_1 B_1, \dots, C_{m+2} = D_{m+2} B_{m+2},$$

where B_i , $i = 1, 2, \dots, m + 2$, is either B or B^* and the chains of the set

$$(D) \quad D_1, D_2, \dots, D_{m+2}$$

are $(m - 1)$ -chains. By the hypothesis of the induction, there can be at most m different $(m - 1)$ -chains in the set (D). Since the members of the set (C) are assumed to be all different, there cannot be three identical $(m - 1)$ -chains in the set (D). It follows that there are at least two pairs in the set (D) such that members of the same pair are identical. We can assume the notation so chosen that

$$\begin{aligned} D_1 &= D_2, & D_3 &= D_4, \\ C_1 &= D_1 B, & C_2 &= D_1 B^*, & C_3 &= D_3 B, & C_4 &= D_3 B^*. \end{aligned}$$

Since the members of the set (C) are all different, it follows that D_1 and D_3 must differ in some cell and can be written in the form

$$D_1 = E_1 B_1 F_1, \quad D_3 = E_3 B_3 F_1,$$

where E_1, F_1, E_3 are chains, while one of the pair B_1, B_3 is B and the other is B^* . The chains

$$B_1F_1B, B_1F_1B^*, B_3F_1B, B_3F_1B^*$$

are subchains of the given set of m -chains. These four chains contain the same number of cells and their b -lengths assume three different values. From this contradiction we infer the truth of the lemma.

LEMMA 3.3. *Corresponding to a given constant $\alpha \geq 0$ the set of Sturmian chains satisfying $(2.11)' \{(2.11)''\}$ contains at most $n + 1$ different n -chains.*

For the set of Sturmian n -chains satisfying $(2.11)' \{(2.11)''\}$ is a set satisfying Condition A and it follows from Lemma 3.2 that there can be at most $n + 1$ different n -chains in the set.

Let T be a Sturmian series and r an integer, positive, negative or zero. The Sturmian series T' which results upon adding r to the index of each cell of T will be said to be *similar* to T . We write $T' \sim T$.

THEOREM 3.5. *A periodic Sturmian series T {beam R } with frequency $\alpha = q/p$, where $(q, p) = 1$, contains $n + 1$ different n -chains if $0 < n < p$ and p different n -chains if $n \geq p$. Two periodic Sturmian series with the same frequency are similar.*

The series T has the minimum cell-period p (cf. Theorem 3.1). It follows that if $n < p$, T must contain at least $n + 1$ different n -chains, for otherwise (cf. SD § 7; the arguments given in SD concern blocks, but similar arguments apply to chains) T would have a cell-period less than p . The b -lengths of the n -chains of T satisfy one of the conditions (2.11) and we infer from Lemma 3.3 that T contains at most $n + 1$ different n -chains. Thus T contains $n + 1$ different n -chains if $0 < n < p$. Since the number of different n -chains of T is a non-decreasing function of n , T contains at least p different n -chains if $n \geq p$. The periodicity of T implies that T contains at most p different n -chains.

Let T and T' be periodic Sturmian series with the same frequency $\alpha = q/p$, $(q, p) = 1$. The b -lengths of the n -chains of T and T' satisfy (3.5) and hence (2.11)'. It follows from Lemma 3.3 that the totality of $(p-1)$ -chains in T and T' form a set containing at most p different $(p-1)$ -chains. But it has been shown that each of the cell-series T and T' contains p different $(p-1)$ -chains and we infer that T and T' contain the same $(p-1)$ -chains. In particular, T and T' contain identical $(p-1)$ -chains and since the b -length of any p -chain of T or T' is q , it follows that T and T' contain

identical p -chains. Since T and T' are periodic with cell-period p , they are similar.

The proof of the theorem is complete.

THEOREM 3.6. *A skew Sturmian series U contains $n + 1$ different n -chains for every positive integer n . Two skew Sturmian series with the same frequency and with critical chains of the same b -length are similar.*

Let n be a positive integer. Since U is not periodic it must contain at least $n + 1$ different n -chains and it follows from Theorem 3.4 and Lemma 3.3 that U contains exactly $n + 1$ different n -chains.

Let U and V be skew Sturmian series with the same frequency $\alpha = q/p$, where $(q, p) = 1$, and with critical p -chains of the same length. The critical p -chains of U and V are either both of b -length $q + 1$ or both of b -length $q - 1$. In the former case U and V satisfy (2.11)' and in the latter (2.11)". It follows from Lemma 3.3 that U and V contain the same n -chains and in particular the same critical p -chains. By virtue of the relation of a skew Sturmian trajectory to its critical chain, as given in Theorem 3.3, we infer that U and V are similar.

Let X be a cell-series and let Y be the cell-series obtained from X by inverting the order of the indices of the cells of X . We term Y the *inverse* of X . If X is a Sturmian series, the inverse of Y satisfies Condition C and hence is Sturmian. If the inverse Y of a Sturmian series X is similar to X , we term X *symmetric*.

COROLLARY. *A skew Sturmian series is symmetric.*

For if X is a skew Sturmian series, its inverse Y is evidently skew Sturmian with frequency equal to that of X and with critical chain of b -length equal to that of the critical chain of X . It follows from Theorem 3.6 that Y is similar to X and hence X is symmetric.

THEOREM 3.7. *Two irrational Sturmian trajectories with the same frequency contain the same $n + 1$ different n -chains for each positive integer n .*

Let T and T' be irrational Sturmian trajectories with the same frequency α . Since neither T nor T' is periodic, each must contain at least $n + 1$ different n -chains for each positive integer n . Since T and T' have the same frequency α the b -lengths of the n -chains of both T and T' satisfy (3.6) and hence (2.11)'. We infer from Lemma 3.3 that the totality of n -chains in both T and T' cannot consist of more than $n + 1$ different n -chains. It follows that T and T' contain the same $n + 1$ different n -chains. The proof of the theorem is complete.

4. Mechanical sequences. Let α be a positive real number and c an arbitrary real number. On the real axis $-\infty < x < +\infty$ we introduce the set of points

$$(4.0) \quad \dots, c - 2\beta, c - \beta, c, c + \beta, c + 2\beta, \dots \quad (\beta = 1/\alpha).$$

We term c the *pole* of this set of points.

Let $T(c, \alpha)$ $\{T'(c, \alpha)\}$ denote the cell-series of the form (2.1) in which the i -th cell B_i contains as many b 's as there are points (4.0) in the interval $i \leq x < i + 1$ $\{i < x \leq i + 1\}$. The b -length of an n -chain of $T(c, \alpha)$ or $T'(c, \alpha)$ is either s_n or $s_n + 1$, where

$$(4.1) \quad n = s_n \beta + r_n \quad (0 \leq r_n < \beta).$$

It follows that $T(c, \alpha)$ and $T'(c, \alpha)$ are Sturmian series. Observe that

$$\alpha = \frac{1}{\beta} = \lim_{n \rightarrow \infty} \left(\frac{s_n}{n} + \frac{r_n}{n\beta} \right) = \lim_{n \rightarrow \infty} \frac{s_n}{n},$$

so that α is the frequency of both $T(c, \alpha)$ and $T'(c, \alpha)$.

When $\alpha = 0$ we understand that (4.0) is a null set of points. The corresponding cell-series contains no b 's and will be denoted by either $T(c, 0)$ or $T'(c, 0)$.

If $c_1 \equiv c_2 \pmod{\beta}$ the corresponding sets (4.0) are identical and $T(c_1, \alpha) \equiv T(c_2, \alpha)$, $T'(c_1, \alpha) \equiv T'(c_2, \alpha)$. The set of points congruent to $x \pmod{\beta}$ will be denoted by $P(x)$. The domain of $P(x)$ will be regarded as a circle Γ . The function $P(x)$ maps the x -axis onto the circle Γ . In this map the image on Γ of a neighborhood of a point c on the x -axis will be regarded as a neighborhood of $P(c)$ on Γ . The circle Γ will be taken in the sense which corresponds locally to the sense of increasing c . The interval PQ on Γ , $P \neq Q$, shall mean the segment of Γ which begins with P and ends with Q , taking Γ in its positive sense, and including P but not Q . When $P = Q$, the interval PQ shall be the whole of Γ . We term Γ the β -circle.

LEMMA 4.1. *If $P(r) \neq P(n+1)$ and $\alpha > 0$, an m -chain $[r, n]$ of $T(c, \alpha)$ is of max or min type respectively according as $P(c)$ is on the interval $P(r)P(n+1)$ or the complementary interval $P(n+1)P(r)$ of the β -circle. If $P(r) = P(n+1)$ all m -chains have the same b -length.*

This follows at once from the conventions upon noting that the type of an m -chain $[r, n]$ of $T(c, \alpha)$ decreases when $P(c)$ leaves the interval $P(r)P(n+1)$ and remains invariant as $P(c)$ varies on this interval, or its complement. In particular, suppose s is an integer between r and $n+1$ inclusive. Suppose $P(r) \neq P(n+1)$. Then $P(s) \neq P(s+1)$ whatever

the integer s . As $P(c)$ enters an interval beginning with $P(s)$, $P(c)$ varying in the positive sense on Γ the cells B_{s-1} and B_s change their types to min and max, respectively.

We define the *alternate interval* PQ , $P \neq Q$, of Γ as the segment of Γ which begins with P and ends with Q , taking Γ in its positive sense, and including Q but not P . When $P = Q$, the alternate interval PQ shall be the whole of Γ . The proof of the following lemma is analogous to that of Lemma 4.1.

LEMMA 4.1'. Read Lemma 4.1 with $T(c, \alpha)$ replaced by $T'(c, \alpha)$ and with the term *interval* replaced by *alternate interval*.

THEOREM 4.1. If α is irrational, two series $T(c, \alpha)$ and $T(a, \alpha)$ $\{T'(c, \alpha)$ and $T'(a, \alpha)\}$ are identical if and only if $c \equiv a \pmod{\beta}$.

If α is irrational, the points $P(n)$, $(n = 1, 2, \dots)$, are everywhere dense on the β -circle Γ . If $c \not\equiv a \pmod{\beta}$, $P(c) \neq P(a)$. There accordingly exist integers r and n with $r < n$ such that $P(c)$ lies on the interval $P(r)P(n+1)$ of Γ while $P(a)$ lies on the complementary interval. It follows from Lemma 4.1 that the chains $[r, n]$ of $T(c, \alpha)$ and $T(a, \alpha)$ are different. If $a \equiv c \pmod{\beta}$, $T(a, \alpha) \equiv T(c, \alpha)$. The proof of the theorem is complete for the Sturmian series $T(c, \alpha)$ and $T(a, \alpha)$.

The proof of the theorem for the series $T'(c, \alpha)$ and $T'(a, \alpha)$ is similar.

The residue intervals. Suppose $\alpha = q/p$ with $q > 0$, $p > 0$, and $(q, p) = 1$. Let m be an arbitrary integer. Then

$$qm = sp + r, \quad 0 \leq r < p,$$

where s and r are integers. Hence

$$m = s \frac{p}{q} + \frac{r}{q}.$$

It follows that the numbers r/q with $r = 0, 1, \dots, p-1$, form a complete set of residues mod β of the rational integers, and the point set $P(n)$ on the β -circle Γ reduces to the set of p points $P(r/q)$. The latter set is identical with the set of points

$$(4.2) \quad P(0), P(1), \dots, P(p-1)$$

since no two integers for which $0 \leq n \leq p-1$ are congruent mod β .

The points (4.2) divide the β -circle Γ into p successive intervals termed *residue intervals* if the initial but not the terminal point is included in an interval, and the *alternate residue intervals* if the terminal but not the initial point is included in an interval.

The following theorem is an easy consequence of Lemmas 4.1 and 4.1'.

THEOREM 4.2. *When α is positive and rational, two cell-series $T(c, \alpha)$ $\{T'(c, \alpha)\}$ are identical if and only if the corresponding points $P(c)$ lie on the same residue interval {alternate residue interval}.*

THEOREM 4.3. *When α is irrational, $T(a, \alpha) \equiv T'(c, \alpha)$ if and only if $a \equiv c \pmod{\beta}$ and $c \not\equiv m \pmod{\beta}$ where m is an integer. If $a \equiv c \equiv m \pmod{\beta}$, m an integer, the corresponding cells of $T(a, \alpha)$ and $T'(c, \alpha)$ are equal except that B_m and B_{m-1} are of max and min type respectively in $T(a, \alpha)$ and of opposite types in $T'(c, \alpha)$.*

If $c \not\equiv m \pmod{\beta}$, where m is an integer, the number of points of the set (4.0) in the interval $i \leq x < i+1$ is identical with the number in the interval $i < x \leq i+1$. Thus the cell B_i of $T(c, \alpha)$ is identical with the cell B_i of $T'(c, \alpha)$ and $T(c, \alpha) \equiv T'(c, \alpha)$. If $a \equiv c \pmod{\beta}$ it follows from Theorem 4.1 that $T(a, \alpha) \equiv T(c, \alpha)$ and hence $T(a, \alpha) \equiv T'(c, \alpha)$.

Conversely, let us assume that $T(a, \alpha) \equiv T'(c, \alpha)$. Since α is irrational the points $P(n)$, $n = 1, 2, \dots$, are everywhere dense on the β -circle Γ and by arguments similar to those given in the proof of Theorem 4.1, it is easily shown that $a \equiv c \pmod{\beta}$. If $c \equiv m \pmod{\beta}$, the interval $m-1 < x \leq m$ contains one more point of the set (4.0) than does the interval $m-1 \leq x < m$, namely the point m . It follows that the cell B_{m-1} is of min type in $T(a, \alpha)$ and of max type in $T'(c, \alpha)$. Similarly, the cell B_m is of max type in $T(a, \alpha)$ and of min type in $T'(c, \alpha)$. Thus if $T(a, \alpha) \equiv T'(c, \alpha)$ we must have $a \equiv c \not\equiv m \pmod{\beta}$.

The second statement of the theorem follows readily.

The following theorem is easily derived with the aid of Theorem 4.2.

THEOREM 4.4. *When α is positive and rational, the cell-series $T(a, \alpha)$ and $T'(c, \alpha)$ are identical if and only if the residue interval in which $P(a)$ lies, coincides, except for end points, with the alternate residue interval in which $P(c)$ lies.*

THEOREM 4.5. *If α is irrational, $T(c, \alpha) \sim T(a, \alpha) \{T'(c, \alpha) \sim T'(a, \alpha)\}$ if and only if $c = a + p\beta + q$, where p and q are integers.*

If $c = a + p\beta + q$ it follows from Theorem 4.1 that $T(c, \alpha) \equiv T(a + q, \alpha)$. But the chain $[r, s]$ of $T(a, \alpha)$ is identical with the chain $[r + q, s + q]$ of $T(a + q, \alpha)$, independently of the values of r and s , and hence $T(a, \alpha) \sim T(a + q, \alpha) \equiv T(c, \alpha)$.

To prove the converse we assume that $T(c, \alpha) \sim T(a, \alpha)$. It follows that there exists an integer q such that the cell B_i of $T(c, \alpha)$ is identical with the cell B_{i-q} of $T(a, \alpha)$ for all integral values of i . The cell B_i of

$T(a+q, \alpha)$ is identical with the cell B_i of $T(c, \alpha)$. Thus $T(c, \alpha) \equiv T(a+q, \alpha)$ and we infer from Theorem 4.1 that $c \equiv a+q, \text{ mod } \beta$, or $c = a + p\beta + q$, where p and q are integers.

A similar proof applies to the pair $T'(c, \alpha)$ and $T'(a, \alpha)$. The proof of the theorem is complete.

THEOREM 4.6. *If α is rational the cell-series $T(c, \alpha)$ and $T'(c, \alpha)$ are periodic and any two of these cell-series are similar, whatever the value of c .*

The periodicity of these cell-series is evident. They are all Sturmian series with the same rational frequency α . It follows from Theorem 3.5 that any two of these cell-series are similar. The proof of the theorem is complete.

If α is irrational and $c \not\equiv m, \text{ mod } \beta$ where m is an integer, the cell-series $T(c, \alpha)$ and $T'(c, \alpha)$ are identical. Thus the class of cell-series $T(c, \alpha)$ corresponding to a given value of α includes most of the cell-series $T'(c, \alpha)$. However, as stated in the following theorem, there are exceptions.

THEOREM 4.7. *If α is irrational and $c \equiv m, \text{ mod } \beta$, where m is an integer, the cell-series $T'(c, \alpha)$ is not similar to any cell-series $T(a, \alpha)$.*

Let us suppose that $T'(c, \alpha)$ and $T(a, \alpha)$ are similar. It follows that there exists an integer q such that $T(a+q, \alpha) \equiv T'(c, \alpha)$. We infer from Theorem 4.3 that $c \equiv m, \text{ mod } \beta$, where m is an integer. From this contradiction we infer the truth of the theorem.

The cell-series $S(m, \alpha)$ and $S'(m, \alpha)$. The preceding cell-series $T(c, \alpha)$ and $T'(c, \alpha)$ include no skew Sturmian series. To obtain such cell-series we introduce new mechanical sequences as follows. Let c be a rational integer m and α positive and rational. In $S(m, \alpha)$ the number of b 's in B_n shall equal the number of points of the set (4.0) on the intervals

$$n < x \leq n+1, \quad m < x < m+1, \quad n \leq x < n+1,$$

according as $n < m$, $n = m$, or $n > m$. In $S'(m, \alpha)$ the number of b 's in B_n shall equal the number of points of the set (4.0) on the intervals

$$n < x \leq n+1, \quad m \leq x \leq m+1, \quad n < x \leq n+1,$$

according as $n < m$, $n = m$, or $n > m$. As a special convention we understand that $S'(m, 0)$ shall consist of null cells except that B_m shall be b . $S(m, 0)$ will not be defined.

THEOREM 4.8. *The cell-sequences $S(m, \alpha)$ and $S'(m, \alpha)$ are skew Sturmian with a critical chain of min and max type respectively. The cell B_m is the initial cell of the critical chain of $S(m, \alpha)$ $\{S'(m, \alpha)\}$.*

In case $\alpha = 0$, all the cells of $S'(m, \alpha)$ are null save B_m and the theorem is obvious. We accordingly assume that $\alpha > 0$.

To establish the theorem it is sufficient to show that $S \{S'\}$ is Sturmian. To that end let θ be the b -length of an n -chain of $S \{S'\}$ whose initial cell is B_r . Then θ is the number of points of the set $(4, 0)$ on an interval $I(S) \{I(S')\}$ of length n beginning at the point $x = r$. Various cases arise according to the nature of n according as I includes one, both, or neither of its end points. We represent n in the form

$$n = s_n \beta + r_n, \quad 0 \leq r_n < \beta,$$

and distinguish between the two following cases.

Case I. $r_n \neq 0$. Here I contains at most one of the points $(4, 0)$ as an end point and $\theta = s_n$ or $s_n + 1$. The comparison Condition C is accordingly satisfied.

Case II. $r_n = 0$. When $r \not\equiv m \pmod{\beta}$, no point of $(4, 0)$ is at an end point of I and $\theta = s_n$. When $r \equiv m \pmod{\beta}$, there are points of $(4, 0)$ at both points of $I(S) \{I(S')\}$. We see that $\theta = s_n$ or $s_n - 1$ in S and s_n or $s_n + 1$ in S' . The Condition C is accordingly satisfied.

Thus S and S' are Sturmian. They have the frequency α . They are not periodic by virtue of the definition of B_m . Moreover S and S' are skew Sturmian with critical p -chain $[m, m + p - 1]$. For the b -lengths of this p -chain in $S(m, \alpha)$ or $S'(m, \alpha)$, respectively, are the number of points of the set $(4, 0)$ on the intervals

$$m < x < m + q\beta, \quad m \leq x \leq m + q\beta, \quad (p = q\beta),$$

and in either case are different from q , the length of every other p -chain. Hence B_m is the initial cell of the critical chain of S or S' .

5. On the representation of Sturmian series by mechanical sequences. The mechanical sequences are Sturmian. We show conversely that any Sturmian series is identical with a properly chosen mechanical sequence.

THEOREM 5.1. *A periodic Sturmian series U with frequency α is identical with $T(c, \alpha)$ for suitable choice of c .*

If $\alpha = 0$ the only symbol appearing in U and $T(c, 0)$ is a and the theorem is evident.

We assume $\alpha = q/p > 0$, where $(q, p) = 1$. According to Theorem 4.6, $T(c, \alpha)$ is a periodic Sturmian series. Since U and $T(c, \alpha)$ are periodic Sturmian series with the same frequency, we infer from Theorem 3.5 that U

and $T(c, \alpha)$ are similar. The series U has the cell-period p and thus there are at most p different Sturmian series which are similar to U and such that no two are identical. According to Theorem 4.2, the cell-series $T(c, \alpha)$ and $T(a, \alpha)$ are identical only if the points $P(c)$ and $P(a)$ lie on the same residue interval. Since there are p residue intervals corresponding to $\alpha = q/p$ it follows that there are p cell-series $T(c, \alpha)$, no two of which are identical. Since all of these cell-series are similar to U , we infer that one of them is identical with U . The proof of the theorem is complete.

THEOREM 5.2. *A skew Sturmian series U with frequency α is identical with one of the cell-series $S(n, \alpha)$ or $S'(n, \alpha)$ for suitable choice of the integer n .*

In the case $\alpha = 0$ the cell-series U is evidently identical with $S'(m, 0)$.

We assume $\alpha = q/p > 0$, where $(q, p) = 1$. Let U be a skew Sturmian series with frequency α , with critical chain of min type and with B_m as the initial cell of its critical chain. Then U and $S(m, \alpha)$ are skew Sturmian series with the same frequency and with critical chains of the same b -lengths. It follows from Theorem 3.6 that U and $S(m, \alpha)$ are similar. According to Theorem 4.8, B_m is the initial cell of the critical chain of $S(m, \alpha)$. It follows from Theorem 3.3 that $S(m, \alpha)$ and U are identical.

Similar arguments show that if the critical chain of U is of max type, U is identical with $S'(m, \alpha)$ for suitable choice of m .

LEMMA 5.1. *An m -chain B whose n -chains satisfy the condition*

$$(5.1) \quad n\alpha - 1 < b_n < n\alpha + 1$$

for some $\alpha \geq 0$ has a copy in the cell-series $T(c, \alpha)$ for each value of c .

The case $\alpha = 0$ is trivial since $b_n = 0$ for each n when $\alpha = 0$.

If α is irrational the cell series $T(c, \alpha)$ contains $m + 1$ different m -chains (cf. Theorem 3.7) whose n -chains satisfy (5.1) (cf. Theorem 3.2). If the n -chains of a set of m -chains satisfy (5.1), they satisfy (2.11)' and are Sturmian. It follows from Lemma 3.3 that the number of different m -chains in such a set cannot exceed $m + 1$. Thus $T(c, \alpha)$ contains all m -chains whose n -chains satisfy (5.1) and in particular $T(c, \alpha)$ contains B .

If $\alpha = q/p \neq 0$, where $(q, p) = 1$, arguments similar to those of the irrational case apply if $m < p$. It follows from (5.1) that all p -chains of B have b -length q ; B is periodic with cell-period p and is completely determined by its initial $(p - 1)$ -chain B^* and the integer q . Since $T(c, q/p)$ contains all the p possible $(p - 1)$ -chains whose n -chains satisfy (5.1) (cf. Theorem 3.5 and proof) it contains B^* . But $T(c, q/p)$ is periodic with cell-

period p and each of its p -chains is of b -length q . It follows that $T(c, q/p)$ contains B .

The proof of the lemma is complete.

THEOREM 5.3. *A Sturmian series U with irrational frequency α is identical with $T(c, \alpha)$ or $T'(c, \alpha)$ for at least one value of c .*

Let $[r, s: a]$ denote the chain $[r, s]$ of $T(a, \alpha)$ and let $[r, s: a]'$ denote the chain $[r, s]$ of $T'(a, \alpha)$. If n is any positive integer it follows from Lemma 5.1 that the chain $[-n, n]$ of U is identical with a chain $[-n + p_n, n + p_n: c]$ of $T(c, \alpha)$ and hence that $[-n, n]$ is identical with the chain $[-n, n: a_n]$ of $T(a_n, \alpha)$ where $a_n = c - p_n$. The points $P(a_n)$ of Γ have a cluster point $P(a)$ and we can assume the sequence n_i , ($i = 1, 2, \dots$), so chosen that the points a_{n_1}, a_{n_2}, \dots vary in one sense on the x -axis and approach $x = a$ as a limit point. With increasing i , the points

$$(5.2) \quad a_{n_i} - m\beta, \dots, a_{n_i} - \beta, a_{n_i}, a_{n_i} + \beta, \dots, a_{n_i} + m\beta,$$

approach the points

$$(5.3) \quad a - m\beta, \dots, a - \beta, a, a + \beta, \dots, a + m\beta,$$

respectively, either from the right or from the left. If $a \equiv k \pmod{\beta}$ for no integer k , no point of the set (5.3) is integral. For a given m and i sufficiently large the chain $[-m, m: a_{n_i}]$ is identical with the chain $[-m, m: a]$. If i is also chosen so large that $m \leq n_i$, the chain $[-m, m]$ of U and the chain $[-m, m: a_{n_i}]$ are identical. It follows that the chain $[-m, m]$ of U is identical with the chain $[-m, m: a]$ of $T(a, \alpha)$ for every positive integer m and hence U is identical with $T(a, \alpha)$.

If $a \equiv k \pmod{\beta}$, where k is an integer, and the points (5.2) approach the points (5.3) from the right, the chains $[-m, m: a_{n_i}]$ and $[-m, m: a]$ are again identical for fixed m and for sufficiently large i . Again U is identical with $T(a, \alpha)$.

If $a \equiv k \pmod{\beta}$, where k is an integer and the points (5.2) approach the points (5.3) from the left it is easily seen that the chain $[-m, m: a_{n_i}]$ and the chain $[-m, m: a]'$ of $T'(a, \alpha)$ are identical for fixed m and for i sufficiently large. In this case the chain $[-m, m]$ of U is identical with the chain $[-m, m: a]'$ of $T'(a, \alpha)$ for every positive integer m . But this implies the identity of U and $T'(a, \alpha)$.

The proof of the theorem is complete.

6. The continuation of Sturmian series. A Sturmian n -chain which appears as a subchain of a Sturmian series T will be said to admit T as a *Sturmian continuation*.

THEOREM 6.1. *Each Sturmian chain x admits aleph continuations.*

By virtue of Theorem 2.3 there exists an open interval of values of α such that the n -chains of x satisfy the relations (5.1). It follows from Lemma 5.1 that for each such value of α there exists a constant c such that $T(c, \alpha)$ contains a copy of x . If $\alpha \neq \alpha'$ the cell series $T(c, \alpha)$ and $T(a, \alpha')$ are not identical so that there are aleph Sturmian continuations of x .

A Sturmian beam whose cells B_i are respectively identical with the cells B'_i with the same index in a Sturmian series T will be said to admit T as a *Sturmian continuation*.

THEOREM 6.2. *Each Sturmian beam R with irrational frequency α admits at least one and at most two Sturmian continuations. In the case where R admits different continuations these continuations are identical respectively with the cell-series $T(m, \alpha)$ and $T'(m, \alpha)$ for a suitable choice of the integer m .*

If the Sturmian beam R with irrational frequency admits a Sturmian continuation, we infer from Theorem 5.3 that this continuation is identical with one of the cell series $T(c, \alpha)$ or $T'(c, \alpha)$ for a suitable choice of c . The Sturmian beam R does not admit two distinct Sturmian continuations of the form $T(c, \alpha)$ and $T(a, \alpha)$. For $T(c, \alpha)$ and $T(a, \alpha)$ would then have a common beam and it would follow as in the proof of Theorem 4.1 that $c \equiv a \pmod{\beta}$, and hence $T(c, \alpha) \equiv T(a, \alpha)$. Similarly R admits at most one Sturmian continuation of the form $T'(c, \alpha)$.

Essentially as in the proof of Theorem 5.3, so here it follows that R possesses at least one continuation of the form $T(c, \alpha)$ or $T'(c, \alpha)$. If $T(c, \alpha)$ and $T'(a, \alpha)$ are different Sturmian continuations of R , it would follow as in the proof of Theorem 4.1 that $a \equiv c \pmod{\beta}$. But since $T(c, \alpha) \not\equiv T'(a, \alpha)$ we conclude from Theorem 4.3 that $c \equiv m \pmod{\beta}$ where m is a suitably chosen integer. But then $T(c, \alpha)$ and $T'(a, \alpha)$ are identical with $T(m, \alpha)$ and $T'(m, \alpha)$ respectively. The proof of the theorem is complete.

THEOREM 6.3. *A non-periodic Sturmian beam R with rational frequency α admits a unique Sturmian continuation.*

Without loss of generality we can assume that R is of the form

$$aB'_k aB'_{k+1} a \cdots$$

If $\alpha = q/p$, where $(q, p) = 1$, it follows from Lemma 3.1 that there exists one and only one p -chain of R whose b -length is different from q . We term this chain the *critical chain* C of R . Exactly as in the proof of Theorem 3.3, so here it follows that the beam following the initial cell B'_m of C and the chain preceding the terminal cell of C are periodic with cell-period p .

The b -length of the n -chains of R satisfy one of the conditions (2.11), say (2.11)". The b -lengths of the n -chains of $S(m, \alpha)$ also satisfy (2.11)" and it follows as in the proof of Theorem 3.6 that the critical p -chains of R and $S(m, \alpha)$ are identical. That their cells have the same indices follows from our choice of m . In view of the relation of $S(m, \alpha)$ to its critical p -chain as disclosed in Theorem 3.3 and the relation of R to its critical p -chain, we can affirm that R appears as a beam of $S(m, \alpha)$ and of no other skew Sturmian series.

In the case where R satisfies (2.11)' similar arguments apply and show that $S'(m, \alpha)$ is the unique continuation of R .

The proof of the theorem is complete.

THEOREM 6.4. *A periodic Sturmian beam R with rational frequency $\alpha = q/p > 0$ admits three dissimilar Sturmian continuations of which one is periodic, and two are skew with critical chains of different type. If $\alpha = 0$, R admits two Sturmian continuations of which one is periodic and the other is skew.*

The proof of Theorem 3.5 shows that if two periodic Sturmian beams have the same frequency $\alpha = q/p$ where $(q, p) = 1$, they have the same cell-period p and contain the same p -chains. If $\alpha > 0$, each of the cell-series $T(c, \alpha)$, $S(m, \alpha)$ and $S'(m, \alpha)$ contains a periodic beam similar to R . If c and m are suitably chosen, each of these cell-series will be a continuation of R . The cell-series $T(c, \alpha)$, $S(m, \alpha)$ and $S'(m, \alpha)$ are dissimilar and it follows from Theorems 3.5 and 3.6 that any other Sturmian series with frequency α is similar to one of these.

If $\alpha = 0$ it is easily seen that $T(c, 0)$ and $S'(m, 0)$, where m is a suitably chosen integer, are the only dissimilar Sturmian continuations of R .

The proof of the theorem is complete.

We distinguish between $T(c, \alpha)$ and $T''(c, \alpha)$ by assigning a *type-index* $+1$ or -1 respectively to these series. We can similarly assign a *type-index* 1 or -1 to the series $S(m, \alpha)$ and $S'(m, \alpha)$ respectively. Thus every Sturmian series T possesses a frequency, at least one pole and a type-index. As we have seen, T admits a mechanical continuation uniquely determined by these numerical characteristics.

In Part II a class of similar Sturmian series will be called a Sturmian trajectory. We note that the members of such a class admit the same numerical characteristics.

II. THE RECURRENCE FUNCTION.

7. Sturmian trajectories and rays. We return to the concept of trajectories and I-trajectories of SD, using the preceding symbols a and b as

generating symbols. Recall that an I-trajectory Λ is an indexed sequence of the form

$$(7.1) \quad \cdots c_{-2}c_{-1}c_0c_1c_2 \cdots$$

in which the symbol c_i is a or b . The class of I-trajectories "similar" to Λ is a trajectory Ω represented by Λ .

Let x be an m -block of Ω . The number of symbols a or b in x will be termed the a -length or b -length respectively of x and written $a(x)$ or $b(x)$. Corresponding to the comparison condition of § 2 we here introduce the following condition.

S. Under Condition S the a -lengths (b -lengths) of two m -blocks with the same m shall differ by at most one.

A trajectory whose blocks satisfy Condition S will be termed a *Sturmian trajectory*.

Prior to the present section we have been considering Sturmian series. Such series are sequences in which the cells are indexed rather than the symbols a and b . They are accordingly logically distinct from indexed trajectories. In the latter the individual symbols a and b are indexed. Each Sturmian series T however defines a trajectory Ω consisting of the symbols a and b appearing in T ordered as in T . We shall say that Ω is *represented* by T .

In any Sturmian trajectory at least one of the symbols a or b appears infinitely many times preceding and following any given symbol. If in particular the symbol a does not so appear, the trajectory T must have one of the two following *special* forms:

$$(7.2) \quad \cdots b b b b b \cdots,$$

$$(7.3) \quad \cdots b b a b b \cdots.$$

These trajectories will be called *b-trajectories*.

The trajectories defined by Sturmian series always include infinitely many a 's and so never include the b -trajectories. More precisely, we have the following theorem.

THEOREM 7.1. *The trajectories defined by Sturmian series satisfy Condition S and include all such trajectories except the b-trajectories.*

Let T be a Sturmian series and let Ω be the trajectory defined by T . Let x and y be arbitrary m -blocks of Ω . Let u be the chain of maximum a -length in x , and v a chain of minimum a -length in T containing y . We see that $a(v) \leq a(y) + 2$, and that $a(u) = a(x)$. If it were true that

$$(7.4) \quad a(y) + 2 \leq a(x),$$

it would follow that $a(v) \leq a(u)$ and there would be a subchain of u with the a -length of v . We could then infer from Condition C that

$$(7.5) \quad b(v) \leq b(u) + 1.$$

From (7.4) and (7.5) we find that

$$a(y) + b(v) \leq a(x) + b(u) - 1.$$

But this is impossible since

$$m \leq a(y) + b(v), \quad a(x) + b(u) \leq m.$$

Relation (7.4) is accordingly false. We infer $|a(x) - a(y)| \leq 1$. Upon recalling that x and y have the same length m we conclude that $|b(x) - b(y)| \leq 1$. The trajectory Ω thus satisfies S.

Conversely, let Ω be an arbitrary Sturmian trajectory which is not a b -trajectory. It is clear that there are no unending sequences of b 's in Ω . The symbols b of Ω can therefore be grouped into maximal blocks of symbols b each preceded and followed by a symbol a . There accordingly exists a cell-sequence T whose symbols a and b appear in T in their order in Ω . It remains to prove that the chains of T satisfy Condition C.

Let x and y be two s -chains of T . If $b(y) < b(x) - 1$, the subblock of x obtained by dropping the two terminal a 's of x would contain a subblock z of the length of y . Then $a(y) - a(z) \geq 2$, contrary to the fact that Ω satisfies Condition S. Hence $b(y) \geq b(x) - 1$. Similarly $b(x) \geq b(y) - 1$. Thus T satisfies Condition C.

We have seen that a Sturmian trajectory Ω which is not a b -trajectory is representable by a Sturmian series T . The n -chains of T will be termed n -chains of Ω . It is clear that the class of n -chains of Ω is independent of the choice of the Sturmian series T representing Ω .

A non-special Sturmian trajectory Ω will be said to have the frequency α of any Sturmian series T representing Ω . It is clear that α is independent of the choice of Sturmian series T representing Ω . A special Sturmian trajectory will be said to have the frequency $\alpha = \infty$.

A Sturmian trajectory Ω defined by an irrational or skew Sturmian series, respectively, will be termed *irrational* or *skew*. The special trajectory (7.3) will also be termed skew Sturmian. Sturmian trajectories defined by periodic Sturmian series are periodic in the sense of SD. They include all periodic Sturmian trajectories except (7.2).

A trajectory Ω is recurrent if corresponding to any positive integer n there exists an integer m such that each m -block of Ω contains a copy of every n -block of Ω . The least such value of m is called the n -th *recurrency index*.

$R(n)$ of Ω and the function $R(n)$ is termed the recurrency function of Ω (cf. SD, p. 827). It is clear that a skew Sturmian trajectory is not recurrent. We shall show that all other Sturmian trajectories are recurrent.

The Sturmian trajectory $\cdots bbb \cdots$ has the recurrency function $R(n) = n$. Any other periodic Sturmian trajectory has a finite rational frequency α and is represented by $T(0, \alpha)$. It is clear that Ω is recurrent. The recurrency function of Ω depends merely on α and will be denoted by $R(n, \alpha)$.

To show that irrational Sturmian trajectories are recurrent we shall need the following lemma.

LEMMA 7.1. *If Ω is a Sturmian trajectory with frequency α and Ω' is a limit trajectory of Ω , then Ω' is a Sturmian trajectory with frequency α .*

Since Ω' is a limit trajectory of Ω , every block of Ω' appears in Ω . Hence Ω' satisfies Condition S and is Sturmian. Every n -chain of Ω' appears in Ω , and since the definition of the frequency α leaves the choice of n -chains arbitrary subject to the condition that n become infinite, it appears that Ω and Ω' have the same frequency. The proof of the lemma is complete.

Now consider a Sturmian trajectory Ω with irrational frequency α . Any limit trajectory Ω' of Ω is Sturmian and has the frequency α . It follows from Theorem 3.7 that Ω and Ω' contain the same chains and hence the same blocks. The permutation number $P(n)$ of Ω (cf. SD, §6) is accordingly identical with that of Ω' , so that Ω is a minimal trajectory. A minimal trajectory is recurrent as stated in Theorem 7.2 of SD. Finally, the recurrency function of Ω depends only on α . For the chains and blocks of Ω are exactly those of $T(0, \alpha)$ so that the recurrency function of Ω is that of $T(0, \alpha)$. We thus have the following theorem.

THEOREM 7.2. *Any Sturmian trajectory with irrational frequency is recurrent with a recurrency function $R(n, \alpha)$ uniquely determined by α .*

8. The derivation of Sturmian trajectories. Let Ω be a Sturmian trajectory represented by a cell-sequence

$$(8.1) \quad \cdots aB_{-1}aB_0aB_1a \cdots$$

Corresponding to Ω we introduce a new trajectory Ω' with an indexed representation

$$(8.2) \quad \cdots c_{-2}c_{-1}c_0c_1c_2 \cdots$$

defined as follows. Let $c_i = a$ if B_i is of minimum type, and let $c_i = b$ if B_i is of maximum type. If all cells B_i are of the same type let $c_i = a$ for all i . The trajectory Ω' will be said to be *derived* from Ω and the I-representation

(8.2) of Ω' will be said to *correspond* to the representation (8.1) of Ω . We proceed with a proof of the following theorem.

THEOREM 8.1. *Let Ω' be a trajectory derived from a recurrent Sturmian trajectory Ω with a frequency α . The trajectory Ω' is Sturmian and has a frequency*

$$(8.3) \quad \alpha' = \frac{\omega}{1 - \omega}$$

where $^3 \omega = \alpha - [\alpha]$.

Let k be the number of b 's in a cell of Ω of minimum type. Suppose Ω represented by (8.1). Let x be an arbitrary n -chain of (8.1) and let y be the corresponding n -block of (8.2). Let b_n denote the b -length of x and let n_a and n_b denote the a -length and b -length respectively of y . Each symbol a in (8.2) corresponds to a cell of (8.1) of b -length k , and each symbol b of (8.2) corresponds to a cell of (8.1) of b -length $k + 1$. Hence

$$b_n = kn_a + (k + 1)n_b.$$

But $n_a + n_b = n$ so that b_n may be given the forms

$$(8.4) \quad b_n = (k + 1)n - n_a = kn + n_b.$$

Since (8.1) is a Sturmian series, b_n varies by at most one for different n -chains x of (8.1). Hence the values of $n_a \{n_b\}$ differ by at most one for different n -blocks y of (8.2). Thus Ω' is Sturmian.

It remains to evaluate the frequency α' of Ω' . First observe that the symbol a occurs infinitely many times preceding and following each symbol of (8.2) since Ω is recurrent. Suppose the block y of (8.2) is an m -chain. Then $m = n_a - 1$ and

$$\alpha' = \lim_{m \rightarrow \infty} \frac{n_b}{m} = \lim_{n \rightarrow \infty} \frac{n_b}{n_a - 1} = \lim_{n \rightarrow \infty} \frac{n_b}{n_a}.$$

Upon making use of (8.4) we see that

$$\alpha' = \lim_{n \rightarrow \infty} \frac{b_n - kn}{(k + 1)n - b_n} = \frac{\alpha - k}{(k + 1) - \alpha}.$$

If α is not an integer it follows from (2.11) that k is the least integer such that

$$\alpha - 1 < k < \alpha + 1.$$

Hence $k = [\alpha]$. If α is an integer each cell of Ω has the b -length α and $k = \alpha = [\alpha]$. Hence (8.3) holds as stated.

³ $[\alpha]$ is the maximum integer not exceeding α .

COROLLARY. *If Ω is periodic or irrational, then Ω' is respectively periodic or irrational.*

For Ω' is recurrent and α' is rational or irrational according as α is rational or irrational.

Let Ω be a recurrent Sturmian trajectory with frequency α and let Ω' be the trajectory derived from Ω . Since Ω is recurrent it has the following property of *chain recurrence*. Given any positive integer n there exists an integer m such that every m -chain of Ω contains a copy of every n -chain of Ω . The least such integer m will be denoted by $\rho(n, \alpha)$ and termed the *chain recurrency function* of Ω . If α' is the frequency of Ω' it is clear that

$$\rho(n, \alpha) = R(n, \alpha') = R\left(n, \frac{\omega}{1 - \omega}\right)$$

where $\omega = \alpha - [\alpha]$. In particular, if $0 \leq \alpha < 1$, then $[\alpha] = 0$ and $\omega = \alpha$ so that

$$(8.5) \quad \rho(n, \alpha) = R\left(n, \frac{\alpha}{1 - \alpha}\right), \quad (0 \leq \alpha < 1).$$

Upon setting

$$\delta = \frac{\alpha}{1 - \alpha}, \quad (0 \leq \alpha < 1)$$

we find that

$$\alpha = \frac{\delta}{1 + \delta}, \quad (0 \leq \delta < \infty)$$

so that (8.5) takes the form

$$(8.6) \quad R(n, \delta) = \rho\left(n, \frac{\delta}{1 + \delta}\right), \quad (0 \leq \delta < \infty).$$

The recurrency function $R(n, \delta)$ of recurrent Sturmian trajectories will accordingly be known once we have determined the chain recurrency function $\rho(n, \alpha)$ for $0 \leq \alpha < 1$. We proceed with a study of chain recurrency functions.

9. The determination of $\rho(n, \alpha)$ in terms of the functions $E(\epsilon, \alpha)$ and $l(n, \alpha)$. The function $\rho(n, \alpha)$ is the chain recurrency function of $T(c, \alpha)$ and is independent of c . We shall suppose that α is irrational inasmuch as the recurrency function of a trajectory with period ω is $\omega + n - 1$ for $n \geq \omega$.

We introduce the points

$$(9.0) \quad P(c+1), P(c+2), \dots, P(c+n), \quad (n \geq 1)$$

on the β -circle Γ . These points are all distinct since α is irrational. They determine n non-overlapping intervals PQ on Γ , where in the special case $n = 1$, $P = Q$ and PQ is the whole of Γ . (For the conventions concerning intervals PQ see § 4.) Let this set of intervals be denoted by $I(c, n, \alpha)$.

Since the set (9.0) can be obtained from a similar set with c replaced by c' by a rotation of Γ into itself, the lengths of the shortest and longest intervals of $I(c, n, \alpha)$ are independent of c and will be denoted by $l(n, \alpha)$ and $L(n, \alpha)$ respectively. When n is infinite in (9.0) the points (9.0) are everywhere dense on Γ and it follows that

$$\lim_{n \rightarrow \infty} l(n, \alpha) = \lim_{n \rightarrow \infty} L(n, \alpha) = 0.$$

Let ϵ be a constant such that $0 < \epsilon \leq \beta$. Let $E(\epsilon, \alpha)$ be the least integer m such that the maximum of the lengths of intervals of $I(c, m, \alpha)$ is at most ϵ . It is clear that $E(\epsilon, \alpha)$ is independent of c in (9.0). We term $E(\epsilon, \alpha)$ the *ergodic function belonging to α* .

Recall that $[r, s; c]$ denotes the chain $[r, s]$ of $T(c, \alpha)$.

LEMMA 9.1. *A set of n -chains*

$$[1, n : a_i], \quad (i = 1, 2, \dots, k),$$

contains all n -chains of $T(c, \alpha)$ if and only if there is a point of the set $P(a_i)$, ($i = 1, \dots, k$), in each of the intervals of the set $I(0, n + 1, \alpha)$.

An arbitrary n -chain $[r, s]$ of $T(c, \alpha)$ is identical with the chain $[1, n]$ of $T(c - r, \alpha)$. Hence the n -chains of $T(c, \alpha)$ are found among the chains $[1, n]$ of $T(a, \alpha)$ for suitable choices of a . Observe that two n -chains $[1, n]$ of $T(a, \alpha)$ and $T(a', \alpha)$ will be identical if corresponding subchains have the same type. Lemma 4.1 gives the conditions under which two such subchains are of the same type, stating these conditions in terms of the intervals of Γ defined by the points

$$P(1), \dots, P(n + 1).$$

Lemma 9.1 follows from Lemma 4.1.

THEOREM 9.1. *The chain recurrency function $\rho(n, \alpha)$ of a recurrent Sturmian trajectory has the value*

$$(9.1) \quad \rho(n, \alpha) = E[l(n + 1, \alpha), \alpha] + n - 1.$$

We shall begin by proving the following:

(a) *If m is an integer which equals the right member of (9.1), then any m -chain x of $T(c, \alpha)$ contains every n -chain of $T(c, \alpha)$.*

The m -chain x is identical with an m -chain $[1, m : a]$ for a suitable choice of a . Since $m \geq n$, the chain $[1, m : a]$ contains the n -chains

$$[1, n : a], [2, n + 1 : a], \dots, [m - n + 1, m : a].$$

These chains are respectively identical with the chains

$$(9.2) \quad [1, n : a], [1, n : a - 1], \dots, [1, n : a - m + n].$$

According to Lemma 9.1, the set (9.2) contains all n -chains of $T(c, \alpha)$ provided there is a point of the set

$$(9.3) \quad P(a), P(a-1), \dots, P(a-m+n)$$

in each of the intervals of $I(0, n+1, \alpha)$.

By hypothesis in (a)

$$m-n+1 = E[l(n+1, \alpha), \alpha]$$

and it follows from the definition of $E(\epsilon, \alpha)$ that the maximum length of the non-overlapping intervals on Γ defined by the points (9.3) is at most $l(n+1, \alpha)$. But the length of the shortest of the intervals $I(0, n+1, \alpha)$ is $l(n+1, \alpha)$ and we conclude that there is a point of the set (9.3) in each of the intervals of $I(0, n+1, \alpha)$. The set (9.2) and hence $[1, m:a]$ and x contain each n -chain of $T(c, \alpha)$. The proof of (a) is complete.

It follows from (a) that when m equals the right member of (9.1) $\rho(n, \alpha) \leq m$. Hence

$$(9.4) \quad \rho(n, \alpha) \leq E[l(n+1, \alpha), \alpha] + n - 1.$$

We shall now suppose that m is an integer such that

$$(9.5) \quad 0 < m - n + 1 < E[l(n+1, \alpha), \alpha]$$

and show that there exists an m -chain of $T(c, \alpha)$ which does not contain every n -chain of $T(c, \alpha)$.

When (9.5) holds there is an interval $\Delta(a)$ of Γ of length $l(n+1, \alpha)$ containing none of the points (9.3). By choosing a properly, the interval $\Delta(a)$ can be brought into coincidence with any given interval of length $l(n+1, \alpha)$ of Γ . There is an interval Δ^* of length $l(n+1, \alpha)$ in the set $I(0, n+1, \alpha)$ and we can assume a so chosen that the interval $\Delta(a)$ coincides with Δ^* . But then the set (9.2) of n -chains does not contain all n -chains of $T(c, \alpha)$ so that the m -chain $[1, m:a]$ does not contain all n -chains of $T(c, \alpha)$. But the chain $[1, m:a]$ is identical with an m -chain of $T(c, \alpha)$, for the cell-series $T(c, \alpha)$ and $T(a, \alpha)$ contain the same set of m -chains. Thus, if (9.5) holds, there is an m -chain of $T(c, \alpha)$ which does not contain all n -chains of $T(c, \alpha)$. We conclude that

$$(9.6) \quad \rho(n, \alpha) \geq E[l(n+1, \alpha), \alpha] + n - 1.$$

The theorem follows from (9.4) and (9.6).

10. The evaluation of the recurrency function of Sturmian trajectories. According to (8.6) the recurrency function $R(n, \alpha)$ of a recurrent Sturmian trajectory with frequency $\alpha > 0$ is the chain recurrency function of a Sturmian trajectory with frequency $\alpha(1+\alpha)^{-1} = \gamma$. As given by (9.1),

the chain recurrency function $\rho(n, \gamma)$ is completely determined by the function $E[l(n+1, \gamma), \gamma]$. We shall show that the latter function bears a simple relation to the denominators D_v of the successive convergents of the continued fraction representing γ .

According to its definition $l(n, \alpha)$, $n \geq 1$, α irrational, is the length of the shortest of the intervals of Γ determined by the points

$$(10.0) \quad P(c+1), P(c+2), \dots, P(c+n).$$

Thus $l(n+1, \alpha)$, where $n \geq 1$, is the length of an interval $P(c+i)P(c+j)$ of Γ where $i \neq j$ and i and j lie between 1 and $n+1$ inclusive. But the length of such an interval is $|s - r\beta| \neq 0$, where $s = |i - j|$ and r is a properly chosen integer. Thus

$$l(n+1, \alpha) = |s - r\beta|$$

where s is an integer such that $0 < s \leq n$. Conversely, if s is an integer such that $0 < s \leq n$ and r is any integer, then either the length of the interval $P(c+1)P(c+s+1)$ or that of its complement on Γ does not exceed $|s - r\beta|$ and hence

$$l(n+1, \alpha) \leq |s - r\beta|.$$

Hence we have the following lemma.

LEMMA 10.0. *The function $l(n+1, \alpha)$ is the least positive value of*

$$|s - r\beta| = \frac{1}{\alpha} |s\alpha - r| \quad (\alpha > 0)$$

as r ranges over all integral values and s assumes the values $1, 2, \dots, n$.

We are concerned with the behavior of $R(n, \alpha)$ for large values of n . If $\alpha = q/p$ where $(p, q) = 1$, the Sturmian trajectory has the period $p+q$ and

$$R(n, \alpha) = p + q + n - 1, \quad n \geq p + q.$$

We turn to the case in which α is irrational. Let

$$\alpha = [b_0, b_1, b_2, \dots]$$

be the development of α as a continued fraction (cf. Perron, p. 39). The integers b_i are uniquely determined by α and with the possible exception of b_0 are positive. The successive convergents A_v/B_v , $v \geq 0$, of α are determined recursively by the formulas

$$(10.1) \quad \begin{cases} A_{-1} = 1, & A_0 = b_0, & A_v = b_v A_{v-1} + A_{v-2}, & (v \geq 1), \\ B_{-1} = 0, & B_0 = 1, & B_v = b_v B_{v-1} + B_{v-2}, & (v \geq 1). \end{cases}$$

As is well known, the integers A_v and B_v are relatively prime and

$$(10.2) \quad 1 = B_0 \leq B_1 < B_2 < \dots$$

Set

$$M_v = B_v - A_v \beta.$$

LEMMA 10.1. *Corresponding to a given irrational α , $l(n+1, \alpha)$ is constant on each interval of the form $B_{v-1} \leq n < B_v$, ($v > 0$) and there has the value $|M_{v-1}|$.*

If n is an integer satisfying the conditions of the lemma and s is an integer such that $0 < s \leq n$, it follows that $0 < s < B_v$. Recall that $(A_v, B_v) = 1$, so that $r/s \neq A_v/B_v$ no matter what the integral choice of r . It follows from a theorem of Lagrange (cf. Perron, p. 52) that

$$(10.3) \quad \frac{1}{\alpha} |s\alpha - r| \geq \frac{1}{\alpha} |B_{v-1}\alpha - A_{v-1}| = |M_{v-1}|.$$

It follows from Lemma 10.0 that

$$l(n+1, \alpha) \geq |M_{v-1}|.$$

The function $l(n+1, \alpha)$ decreases monotonically with n so that for $n \geq B_{v-1}$,

$$l(n+1, \alpha) \leq l(B_{v-1}+1, \alpha).$$

By virtue of Lemma 10.0, $l(B_{v-1}+1, \alpha)$ does not exceed the value of $|s - r\beta|$ when $s = B_{v-1}$ and $r = A_{v-1}$ so that

$$l(n+1, \alpha) \leq |B_{v-1} - A_{v-1}\beta| = |M_{v-1}|.$$

The proof of the lemma is complete.

Recall that $L(n, \alpha)$ is the maximum length of the non-overlapping intervals of Γ determined by the points (10.0), and that it is independent of c in (10.0).

LEMMA 10.2. *For α irrational and $v > 0$*

$$(10.5) \quad L(B_{v-1} + B_v, \alpha) \leq |M_{v-1}|,$$

$$(10.5)' \quad L(B_{v-1} + B_v - 1, \alpha) > |M_{v-1}|,$$

except when $v = 1$ and $B_0 = B_1$.

The left member of (10.5) is the maximum length of the non-overlapping intervals of Γ determined by the points

$$(10.6) \quad P(1), P(2), \dots, P(B_{v-1} + B_v).$$

To prove (10.5) it is sufficient to show that each point of the set (10.6) is followed on Γ by a point of (10.6) at a distance not exceeding $|M_{v-1}|$. We distinguish two cases according to the sign of M_{v-1} .

Case I. $M_{v-1} > 0$. It follows from Perron, p. 42, that

$$0 < M_{v-1} = \frac{1}{\alpha} (B_{v-1}\alpha - A_{v-1}) < \frac{1}{\alpha} |B_{v-1}\alpha - A_{v-1}| = \beta.$$

On the axis of reals the point $B_{v-1} + i$ follows the point i at a distance B_{v-1} . But $B_{v-1} \equiv M_{v-1} \pmod{\beta}$ and M_{v-1} lies between 0 and β so that $P(B_{v-1} + i)$ follows $P(i)$ on Γ at a distance equal to M_{v-1} . Since $M_{v-1} > 0$, it follows from Perron, p. 42, that M_v lies between 0 and $-\beta$ so that the point $P(i)$ follows $P(B_v + i)$ on Γ at a distance equal to $|M_v| < M_{v-1}$. Thus each point of the set (10.6) is followed on Γ by a point of (10.6) at a distance not exceeding M_{v-1} . The proof of (10.5) is complete in Case I.

Case II. $M_{v-1} < 0$. Arguments similar to those given in Case I show that each point of the set (10.6) is preceded (and hence followed) by a point of (10.6) at a distance not exceeding $|M_{v-1}|$.

The proof of (10.5) is complete.

We turn to the proof of (10.5)'. It follows from Lemma 10.1 that

$$l(B_v, \alpha) = |M_{v-1}| = l(B_{v-1} + 1, \alpha)$$

so that there is no point of the set

$$(10.7) \quad P(1), P(2), \dots, P(B_v - 1)$$

in the interior of an interval I of length $2|M_{v-1}|$ with $P(B_v)$ as midpoint. Both end points of I cannot belong to the set (10.7). For if this were the case and these end points were $P(i)$ and $P(j)$ where $1 \leq i < j \leq B_v - 1$, then

$$B_v - i \equiv j - B_v \pmod{\beta}.$$

But this is impossible if α and hence β is irrational. Thus it is clear that there is an interval I^* of Γ of length exceeding $|M_{v-1}|$, with $P(B_v)$ as one of its end points, and with no point of the set (10.7) or $P(B_v)$ in its interior.

Consider the set

$$(10.8) \quad P(B_v), P(B_v + 1), \dots, P(B_{v-1} + B_v - 1).$$

There are B_{v-1} points in this set and if $v = 1$, or if $v = 2$ and $B_0 = B_1 = 1$, the set consists of the single point $P(B_v)$ which is not in the interior of I^* . In any other case it follows from Lemma 10.1 that the shortest distance on Γ between points of the set (10.8) is $|M_{v-2}| > |M_{v-1}|$. But then a suitably chosen subinterval I^{**} of I^* contains no points of the set (10.8) in its interior and has a length exceeding $|M_{v-1}|$. There are no points of the sets (10.7) or (10.8) in I^{**} . This implies (10.5)'.

The proof of the lemma is complete.

LEMMA 10.3. If $v > 0$ and $B_{v-1} \leq n < B_v$,

$$(10.9) \quad E\{l(n + 1, \alpha), \alpha\} = B_{v-1} + B_v.$$

For if $B_{v-1} \leq n < B_v$, it follows from Lemma 10.1 that

$$l(n+1, \alpha) = |M_{v-1}|.$$

According to (10.5)' there exists an interval of Γ of length exceeding $|M_{v-1}|$ which contains none of the points

$$(10.10) \quad P(1), P(2), \dots, P(B_{v-1} + B_v - 1)$$

and hence

$$(10.11) \quad E\{l(n+1, \alpha), \alpha\} > B_{v-1} + B_v - 1.$$

It follows from (10.5) that every interval of Γ of length $|M_{v-1}|$ contains a point of the set (10.6) and thus

$$(10.12) \quad E\{l(n+1, \alpha), \alpha\} \leq B_{v-1} + B_v.$$

The equality (10.9) is implied by (10.11) and (10.12).

Recall that $\gamma = \alpha(1 + \alpha)^{-1}$. Let $\gamma = [d_0, d_1, d_2, \dots]$ be the continued fraction representation of γ and let C_v/D_v be the corresponding v -th convergent.

THEOREM 10.1. *The recurrency function $R(n, \alpha)$ increases by unity when n increases from $n-1$ to n except when n is a denominator D_v of $\gamma = \alpha(1 + \alpha)^{-1}$. For these exceptional values of n ,*

$$(10.13) \quad R(D_v, \alpha) = D_{v+1} + 2D_v - 1 \quad (v \geq 0)$$

starting with D_1 when $D_0 = D_1$. $R(n, \alpha)$ is thereby uniquely determined for all positive integers n .

According to (8.6) and (9.1)

$$(10.14) \quad R(n, \alpha) = \rho(n, \gamma) = E\{l(n+1, \gamma), \gamma\} + n - 1.$$

If $D_{v-1} \leq n < D_v$, it follows from (10.14) and (10.9) that

$$(10.15) \quad R(n, \alpha) = D_v + D_{v-1} + n - 1$$

and consequently if $D_{v-1} < n < D_v$,

$$(10.16) \quad R(n, \alpha) = D_v + D_{v-1} + n - 1 = R(n-1, \alpha) + 1.$$

But when α and hence γ is irrational, each positive integer n not a denominator of γ lies between two successive denominators of γ . Thus (10.16) holds if n is not a denominator of γ . Upon setting $n = D_{v-1}$ in (10.15) we find that

$$R(D_{v-1}, \alpha) = D_v + 2D_{v-1} - 1$$

excepting the case where $D_{v-1} = D_0 = D_1$.

We infer the truth of the theorem.

THEOREM 10.2. *For irrational α the recurrency functions $R(n, \alpha)$ and $R(n, 1/\alpha)$ are identical; conversely if $R(n, \alpha)$ and $R(n, \alpha')$ are equal for all values of n , either $\alpha' = \alpha$ or $\alpha' = \alpha^{-1}$.*

Let Ω be the Sturmian trajectory defined by $T(c, \alpha)$. The function $R(n, \alpha)$ is the recurrency function of Ω . The trajectory Ω^* obtained from $T(c, \alpha)$ by replacing a by b and b by a has the frequency α^{-1} . But the definition of the recurrency function of a Sturmian trajectory is symmetric with respect to a and b and it follows that the recurrency function $R(n, \alpha^{-1})$ of Ω^* is identical with $R(n, \alpha)$.

To prove the converse let us assume that $R(n, \alpha)$ and $R(n, \alpha')$ are equal for all values of n . It follows from Theorem 10.1 that the denominators of $\alpha(1 + \alpha)^{-1}$ and of $\alpha'(1 + \alpha')^{-1}$ are identical as sets of numbers. Let D_ν and D'_ν be respectively the ν -th denominators of $\alpha(1 + \alpha)^{-1}$ and of $\alpha'(1 + \alpha')^{-1}$ with $\nu \geq 0$. We distinguish between four cases:

Case I. $D_0 < D_1, D'_0 < D'_1$.

Case II. $D_0 = D_1, D'_0 = D'_1$.

Case III. $D_0 = D_1, D'_0 < D'_1$.

Case IV. $D_0 < D_1, D'_0 = D'_1$.

The values assumed by the denominators D_ν form a set of numbers identical with the set of numbers assumed by the denominators D'_ν . In Cases I and II it follows that $D_i = D'_i$ for all admissible values of i . But this implies that the continued fraction developments of $\alpha(1 + \alpha)^{-1}$ and $\alpha'(1 + \alpha')^{-1}$ are identical. Consequently these members are equal and $\alpha = \alpha'$.

In Case III it is clear that $D_{i+1} = D'_i$ for each non-negative integer i . It follows that

$$\frac{\alpha}{1 + \alpha} = [0, 1, d_2, d_3, \dots],$$

$$\frac{\alpha'}{1 + \alpha'} = [0, d_2 + 1, d_3, d_4, \dots].$$

It is easily shown that $\alpha\alpha' = 1$. The proof in Case IV is similar to the proof in Case III.

11. The asymptotic behavior of $R(n, \alpha)$. We continue with the case of an irrational frequency α . The constant $\gamma = \alpha(1 + \alpha)^{-1}$ is then irrational and the denominators of γ form an infinite sequence

$$(11.1) \quad 1 = D_0 \leq D_1 < D_2 < \dots$$

Given a positive integer n there exists a unique non-negative integer ν such that $D_\nu \leq n < D_{\nu+1}$ and according to (10.15), for these values of n

$$(11.2) \quad R(n, \alpha) = D_{\nu+1} + D_\nu + n - 1, \quad (D_\nu \leq n < D_{\nu+1}).$$

This implies that

$$(11.3) \quad 2 + \frac{D_\nu}{D_{\nu+1} - 1} \leq \frac{R(n, \alpha)}{n} \leq 2 + \frac{D_{\nu+1} - 1}{D_\nu}, \quad (D_\nu \leq n < D_{\nu+1}),$$

where the equalities on the left and right hold respectively when $n = D_{v+1} - 1$ and $n = D_v$. Let

$$\limsup_{v \rightarrow \infty} \frac{D_{v+1}}{D_v} = \lambda(\alpha) \geq 1,$$

and recall that $\lambda(\alpha)$ may be infinite. It follows from (11.3) that

$$(11.4) \quad \limsup_{n \rightarrow \infty} \frac{R(n, \alpha)}{n} = 2 + \lambda(\alpha),$$

$$(11.5) \quad \liminf_{n \rightarrow \infty} \frac{R(n, \alpha)}{n} = 2 + \lambda^{-1}(\alpha),$$

understanding that $\lambda^{-1}(\alpha) = 0$ if $\lambda(\alpha) = \infty$. If $\lambda(\alpha)$ is finite the closed interval

$$2 + \lambda^{-1}(\alpha) \leq x \leq 2 + \lambda(\alpha)$$

will be called the *limit range* of $R(n, \alpha)/n$ corresponding to α . If $\lambda(\alpha)$ is infinite this limit range shall be $2 \leq x < \infty$.

THEOREM 11.1. *The limit range corresponding to any irrational α is of length at least 1 and on this range the numbers $R(n, \alpha)/n$, ($n = 1, 2, \dots$), are everywhere dense.*

It follows from (10.1) that

$$\frac{D_{v+1}}{D_v} - \frac{D_{v-1}}{D_v} = d_{v+1} \geq 1 \quad (v \geq 0)$$

and hence if $\lambda(\alpha)$ is finite

$$\limsup_{v \rightarrow \infty} \frac{D_{v+1}}{D_v} - \liminf_{v \rightarrow \infty} \frac{D_{v-1}}{D_v} = \lambda(\alpha) - \frac{1}{\lambda(\alpha)} \geq 1.$$

The length of the limit range in this case is least 1. If $\lambda(\alpha) = \infty$, the length of the limit range is evidently infinite.

To prove that the set $R(n, \alpha)/n$ is everywhere dense on the limit range, let x be any point in the (open) interior of the limit range. It follows from (11.3) that there exist arbitrarily large values of v such that

$$2 + \frac{D_v}{D_{v+1} - 1} < x < 2 + \frac{D_{v+1} - 1}{D_v}.$$

According to (11.2), if n is an integer such that $D_v \leq n < n + 1 < D_{v+1}$, then

$$\frac{R(n+1, \alpha)}{n+1} < \frac{R(n, \alpha)}{n}$$

and thus if n is a properly chosen integer

$$\frac{R(n+1, \alpha)}{n+1} \leq x \leq \frac{R(n, \alpha)}{n}, \quad (D_v \leq n < n + 1 < D_{v+1}).$$

Since $R(n+1, \alpha)$ exceeds $R(n, \alpha)$ by 1, the length of this interval is

$$\frac{R(n, \alpha)}{n} - \frac{R(n+1, \alpha)}{n+1} = \frac{R(n+1, \alpha) - (n+1)}{n(n+1)} \leq \frac{x-1}{n}.$$

But n becomes infinite with ν so that the length of this interval approaches zero. This implies that the set $R(n, \alpha)/n$ is everywhere dense on the limit range, and the proof of the theorem is complete.

The Sturmian trajectories thus yield no example of a non-periodic Sturmian trajectory with recurrency function $R(n)$ such that $R(n)/n$ has a finite limit as n becomes infinite. Whether or not there exist more general non-periodic recurrent trajectories such that this limit exists is at the present unknown.

The numbers α and α' will be said to be *equivalent* if there exist integers a, b, c, d with $ad - bc = \pm 1$ such that

$$\alpha' = \frac{a\alpha + b}{c\alpha + d}.$$

THEOREM 11.2. *The limit ranges corresponding to equivalent irrational values of α are identical.*

Let α and α' be equivalent irrational numbers. Then $\alpha(1+\alpha)^{-1}$ and $\alpha'(1+\alpha')^{-1}$ are irrational and can be represented by continued fractions

$$\begin{aligned}\alpha(1+\alpha)^{-1} &= [d_0, d_1, d_2, \dots], \\ \alpha'(1+\alpha')^{-1} &= [d'_0, d'_1, d'_2, \dots].\end{aligned}$$

It is readily shown that the equivalence of α and α' implies the equivalence of $\alpha(1+\alpha)^{-1}$ and $\alpha'(1+\alpha')^{-1}$ and consequently (Perron, p. 65) there exist integers k and j such that

$$d_{k+i} = d'_{j+i}, \quad i \geq 1.$$

Let D_ν be the denominator of the ν -th convergent of $\alpha(1+\alpha)^{-1}$ and let D'_ν be the denominator of the ν -th convergent of $\alpha'(1+\alpha')^{-1}$. Then (cf. Perron, p. 32)

$$\begin{aligned}\frac{D_{k+i}}{D_{k+i-1}} &= [d_{k+i}, d_{k+i-1}, \dots, d_{k+1}, d_k, \dots, d_1], \\ \frac{D_{j+i}}{D_{j+i-1}} &= [d'_{j+i}, d'_{j+i-1}, \dots, d'_{j+1}, d'_j, \dots, d'_1] \\ &= [d_{k+i}, d_{k+i-1}, \dots, d_{k+1}, d'_j, \dots, d'_1].\end{aligned}$$

It follows that

$$\lim_{i \rightarrow \infty} \left(\frac{D_{k+i}}{D_{k+i-1}} - \frac{D'_{j+i}}{D'_{j+i-1}} \right) = 0$$

and hence

$$\lambda(\alpha) = \lambda(\alpha').$$

The theorem follows directly.

THEOREM 11.3. *The limit range corresponding to $(\sqrt{5} + 1)/2$ is the interval*

$$(11.6) \quad \frac{3 + \sqrt{5}}{2} \leq x \leq \frac{5 + \sqrt{5}}{2}.$$

The limit range corresponding to any irrational α which is not equivalent to $(\sqrt{5} + 1)/2$ contains the interval (11.6) in its interior.

If $\alpha = (\sqrt{5} + 1)/2$,

$$\gamma = \frac{\alpha}{1 + \alpha} = \frac{\sqrt{5} - 1}{2} = [0, 1, 1, 1, \dots] = [d_0, d_1, d_2, \dots].$$

But then

$$\frac{D_{v+1}}{D_v} = [d_{v+1}, d_v, \dots, d_2, d_1] = [1, 1, 1, \dots, 1, 1]$$

and hence

$$\lambda(\alpha) = \lim_{v \rightarrow \infty} \frac{D_{v+1}}{D_v} = [1, 1, 1, \dots] = \frac{\sqrt{5} + 1}{2}.$$

Thus the limit range corresponding to $\alpha = (\sqrt{5} + 1)/2$ is

$$\frac{3 + \sqrt{5}}{2} = 2 + \lambda^{-1}(\alpha) \leq x \leq 2 + \lambda(\alpha) = \frac{5 + \sqrt{5}}{2}.$$

It follows from Theorem 11.2 that the limit range corresponding to any α which is equivalent to $(\sqrt{5} + 1)/2$ is also the interval (11.6).

Suppose α is not equivalent to $(\sqrt{5} + 1)/2$. Let $[d_0, d_1, \dots]$ represent the new value of γ and let C_v/D_v be the v -th convergent of γ . It follows from Perron, p. 65, that there exists no integer k such that

$$1 = d_k = d_{k+1} = \dots$$

Consequently there exists an infinite sequence $v_1 < v_2 < \dots$ such that $d_{v_i+1} \geq 2$, ($i = 1, 2, \dots$). But then

$$\frac{D_{v_i+1}}{D_{v_i}} = [d_{v_i+1}, d_{v_i}, \dots, d_2, d_1] \geq 2,$$

and hence

$$\lambda(\alpha) = \limsup_{v \rightarrow \infty} \frac{D_{v+1}}{D_v} \geq 2.$$

It follows that the limit range of α contains the interval

$$\frac{5}{2} \leq x \leq 4$$

which includes (11.6) in its interior.

The proof of the theorem is complete.

THEOREM 11.4. *If $\alpha = (\sqrt{5} + 1)/2$,*

$$\frac{3 + \sqrt{5}}{2} < \frac{R(n, \alpha)}{n} < \frac{5 + \sqrt{5}}{2}.$$

For the given value of α

$$\frac{\alpha}{1 + \alpha} = \frac{1}{\alpha} = \alpha - 1 = \frac{\sqrt{5} - 1}{2} = [0, 1, 1, 1, \dots].$$

The relations (10.1) corresponding to the convergents C_v/D_v of $\alpha(1 + \alpha)^{-1}$ become

$$(11.7) \quad \begin{cases} C_{-1} = 1, & C_0 = 0, & C_{v+1} = C_v + C_{v-1}, \\ D_{-1} = 0, & D_0 = 1, & D_{v+1} = D_v + D_{v-1}, \end{cases} \quad \begin{matrix} (v \geq 0), \\ (v \geq 0). \end{matrix}$$

It is an easy consequence of these relations that

$$(11.8) \quad C_v = D_{v-1}, \quad (v \geq 0).$$

Since $D_0 = D_1 = 1$ each positive integer n lies on an interval of the form

$$D_v \leq n < D_{v+1}, \quad (v > 0).$$

It follows from (11.3), (11.7) and (11.8) that

$$(11.9) \quad \frac{R(n, \alpha)}{n} \leq 2 + \frac{D_{v+1} - 1}{D_v} = 3 + \frac{C_v - 1}{D_v} = 2 + \alpha + \frac{C_v}{D_v} - \frac{1}{\alpha} - \frac{1}{D_v}.$$

By virtue of Perron, p. 43, and the relation $\alpha(1 + \alpha)^{-1} = \alpha^{-1}$,

$$\frac{1}{D_v} > \frac{1}{D_v D_{v+1}} \geq \left| \frac{C_v}{D_v} - \frac{\alpha}{1 + \alpha} \right| = \left| \frac{C_v}{D_v} - \frac{1}{\alpha} \right|, \quad (v > 0).$$

It follows that

$$\frac{C_v}{D_v} - \frac{1}{\alpha} - \frac{1}{D_v} < 0, \quad (v > 0),$$

and upon using (11.9) that

$$\frac{R(n, \alpha)}{n} < 2 + \alpha = \frac{5 + \sqrt{5}}{2}.$$

Similarly it follows from (11.3) and (11.8) that

$$\frac{R(n, \alpha)}{n} \geq 2 + \frac{D_v}{D_{v+1} - 1} = 1 + \alpha + \frac{C_{v+1}}{D_{v+1}} - \frac{1}{\alpha} + \frac{C_{v+1}}{(D_{v+1} - 1)D_{v+1}}.$$

But

$$\left| \frac{C_{v+1}}{D_{v+1}} - \frac{1}{\alpha} \right| \leq \frac{1}{D_{v+1}D_{v+2}} < \frac{C_{v+1}}{(D_{v+1} - 1)D_{v+1}}, \quad (v > 0),$$

and hence

$$\frac{R(n, \alpha)}{n} > 1 + \alpha = \frac{3 + \sqrt{5}}{2}.$$

The proof of the theorem is complete.

It is clear that $R(n, \alpha)$ becomes infinite with n . For an irrational value

of α it follows from SD, p. 830, that $R(n, \alpha) \geq 2n$ without exception. The following theorem is concerned with a more precise description of the manner in which $R(n, \alpha)$ becomes infinite with n . It is natural in analysis of this character to be concerned with $R(n, \alpha)$ for "almost all values" of α rather than with the function $R(n, \alpha)$ for individual values of α .

Let $\phi(x)$ be a function which is positive and non-decreasing for $x \geq 0$ and such that $\lim_{x \rightarrow \infty} \phi(x) = +\infty$.

THEOREM 11.5. *For almost all α*

$$\limsup_{n \rightarrow \infty} \frac{R(n, \alpha)}{n\phi(\log n)}$$

is finite or infinite according as the series

$$\sum_{n=0}^{\infty} \frac{1}{\phi(n)}$$

is convergent or divergent.

Since the rational values of α form a set of measure zero, it is sufficient to prove the theorem for almost all irrational values of α .

Let

$$S(\alpha) = \limsup_{n \rightarrow \infty} \frac{R(n, \alpha)}{n\phi(\log n)}.$$

If α is irrational, it follows from (11.3) that

$$S(\alpha) = \limsup_{\nu \rightarrow \infty} \frac{R(D_\nu, \alpha)}{D_\nu \phi(\log D_\nu)} = \limsup_{\nu \rightarrow \infty} \frac{D_{\nu+1}}{D_\nu \phi(\log D_\nu)},$$

where D_ν is the ν -th denominator of $\alpha(1+\alpha)^{-1} = \gamma$. With the aid of (10.1) we infer that

$$(11.10) \quad S(\alpha) = \limsup_{\nu \rightarrow \infty} \frac{d_{\nu+1}}{\phi(\log D_\nu)}$$

where the integers d_ν are those appearing in the continued fraction $[d_0, d_1, \dots]$ representing γ .

But it is known (cf. Lévy, p. 289) that for almost all values of γ and hence for almost all irrational values of α

$$(11.11) \quad \lim_{\nu \rightarrow \infty} \sqrt[\nu]{D_\nu} = K,$$

where K is an absolute constant such that $3 < K < 4$. On applying (11.11) to (11.10) we infer that for almost all irrational values of

$$(11.12) \quad \limsup_{\nu \rightarrow \infty} \frac{d_{\nu+1}}{\phi(\nu \log 4)} \leq S(\alpha) \leq \limsup_{\nu \rightarrow \infty} \frac{d_{\nu+1}}{\phi(\nu \log 3)}.$$

According to a theorem of Borel and Bernstein (cf. Koksma, p. 46,

Theorem 14), if $\psi(x)$, $x \geq 1$, is a function which is positive and non-decreasing, the relation

$$b_\nu = O(\psi(\nu))$$

is true for almost all γ or false for almost all γ according as the series

$$\sum_{\nu=1}^{\infty} \frac{1}{\psi(\nu)}$$

is convergent or divergent. If the series

$$(11.13) \quad \sum_{\nu=0}^{\infty} \frac{1}{\phi(\nu)}$$

is convergent, it follows that the series

$$\sum_{\nu=0}^{\infty} \frac{1}{\phi(\nu \log 3)}$$

is convergent. On setting $\psi(x+1) = \phi(x \log 3)$, $x \geq 0$, we infer from the Borel-Bernstein theorem that for almost all γ

$$d_{\nu+1} = O(\phi(\nu \log 3)).$$

It follows from (11.12) that $S(\alpha)$ is finite for almost all α .

If the series (11.13) is divergent, it is easily shown that due to the hypotheses on $\phi(x)$ the series

$$\sum_{\nu=0}^{\infty} \frac{1}{\phi(\nu \log 4)}$$

is divergent. On setting $\psi(x+1) = \phi(x \log 4)$, $x \geq 0$, we infer from the Borel-Bernstein theorem that for almost all γ the relation

$$d_{\nu+1} = O(\phi(\nu \log 4))$$

does not hold. It follows from (11.12) that for almost all α , $S(\alpha)$ is not finite.

The proof of the theorem is complete.

If we set $\phi(x) = x$, $x \geq 1$, and $\phi(x) = 1$, $0 \leq x \leq 1$, the following corollary is an immediate consequence of Theorem 11.5.

COROLLARY 11.1. *For almost all α*

$$\limsup \frac{R(n, \alpha)}{n \log n} = +\infty.$$

This implies in particular that for almost all α

$$\limsup_{n \rightarrow \infty} \frac{R(n, \alpha)}{n} = +\infty.$$

If $a > 1$ and we set $\phi(x) = x^a$, $x \geq 1$, $\phi(x) = 1$, $0 \leq x \leq 1$, in Theorem 11.5 we obtain the following corollary.

COROLLARY 11.2. If $a > 1$,

$$\limsup_{n \rightarrow \infty} \frac{R(n, \alpha)}{n(\log n)^a} = 0,$$

except for a set of values of α of measure zero.

We have excluded the case in which α is rational because for $\alpha = q/p$ with $(q, p) = 1$ the periodic Sturmian trajectory with frequency α has the period $\omega = p + q$ and

$$(11.14) \quad R(n, \alpha) = \omega + n - 1, \quad n \geq \omega.$$

If α is an integer, (11.14) holds for $n > 0$. If α is rational but not an integer, (11.14) does not hold for all $n < \omega$, as we shall see. However, the values of $R(n, \alpha)$ for $n < \omega$ can be determined by methods similar to those applicable to the case when α is irrational.

If $\alpha = q/p$, where $(q, p) = 1$ and α is not an integer, and if we set

$$\gamma = \frac{\alpha}{1 + \alpha} = \frac{q}{p + q},$$

γ is not an integer and admits a unique representation in the form of a continued fraction (cf. Perron, p. 30)

$$\gamma = [d_0, d_1, \dots, d_\mu], \quad \mu \geq 1, \quad d_\mu \geq 2.$$

The recursion formulas (10.1) determining the successive convergents C_ν/D_ν of γ are valid for $\nu \leq \mu$. We state the following theorem without proof.

THEOREM 11.6. If $\alpha = q/p$ where $(q, p) = 1$ and $p \neq 1$, Theorem 10.1 holds for $n < D_{\mu-1}$. For $n \geq D_{\mu-1}$

$$R(n, \alpha) = p + q - 1.$$

By means of this theorem it is easily shown that Theorem 10.2 is valid for positive rational as well as irrational values of α .

12. Sturmian sequences in differential equation theory. We are concerned here with linear homogeneous second order differential equations with coefficients which are continuous in the independent variable x . We shall make use of the important canonical form

$$(12.1) \quad y'' + \phi(x)y = 0.$$

We assume that $\phi(x)$ has the period 1. Corresponding to an arbitrary solution $u(x)$ of (12.1) with $u \not\equiv 0$, let $T(u)$ and $T'(u)$ be respectively cell-series

$$\dots aB_{-1}aB_0aB_1a \dots$$

in which B_n is the number of zeros of u on the intervals

$$n \leq x < n + 1, \quad n < x \leq n + 1.$$

It follows from the well-known Sturmian separation theorem that $T(u)$ and $T'(u)$ are Sturmian in the sense of §2. Moreover the frequency α of $T(u)$ or of $T'(u)$ depends only on $\phi(x)$ and not upon the choice of the solution u . We may refer to α as the frequency of (12.1).

The cell-series $T(u)$ and $T'(u)$ include all of the types which we met in the general study. Consider for example the equation

$$y'' + a^2y = 0$$

where a is a positive constant. The corresponding cell-series $T(u)$ and $T'(u)$ have the frequency $a/\pi = \alpha$ and, as is easily seen, include all of the types $T(c, \alpha)$ and $T'(c, \alpha)$, respectively, for suitable choices of a and of u . When $a = 0$ we have the solution x . The series $T(x)$ is skew Sturmian of the form $S'(0, 0)$. To obtain skew Sturmian series $T(u)$ more general in form it is necessary to go somewhat deeper.

We recall a few facts in the classical theory of differential equations of the type (12.1). Let $y(x)$ and $w(x)$ be solutions of (12.1). Keeping x real we admit solutions of (12.1) of the form $Ay(x) + Bw(x)$, where A and B are complex constants. Let p be an arbitrary positive integer. As is well known, there exists at least one solution $u(x)$ of (12.1) such that

$$(12.2) \quad u(x + p) \equiv \rho u(x), \quad (\rho \neq 0),$$

where ρ is a real or complex constant. We term ρ a *characteristic root of index p* . The roots ρ satisfy a quadratic equation, the product of whose roots is 1. There are two principal cases according as the roots ρ are real and positive, or not real and complex. There is also the degenerate case in which the roots are equal. The equation (12.1) possesses a canonical pair of independent solutions whose properties depend upon the classification of the roots ρ . It would not be difficult to show the precise connection between these canonical forms and the types of trajectories $T(u)$ and $T'(u)$ defined by the solutions of (12.1). We shall not go into details beyond proving the following theorem.

THEOREM 12.1. *In case the differential equation (12.1) possesses two real positive unequal characteristic roots of index p , then for a suitable choice of the origin and of the solution $u(x)$, the series $T(u)$ and $T'(u)$ are skew Sturmian trajectories with a frequency of the form q/p .*

Let c be one of the roots of index p . The reciprocal of c is another such root. We set $a = p^{-1} \log c$. It is easy to prove that there are two independent solutions of (12.1) of the form

$$(12.3) \quad \begin{cases} y(x) = e^{ax}A(x), \\ w(x) = e^{-ax}B(x), \end{cases}$$

where $A(x)$ and $B(x)$ have the period p . Moreover

$$\begin{aligned}y(x+p) &\equiv cy(x), \\w(x+p) &\equiv \frac{1}{c}w(x),\end{aligned}$$

and the only solutions $u(x)$ of (12.1) which satisfy a relation of the form (12.2) are the constant multiples of $y(x)$ and $w(x)$.

Suppose that $A(x)$ vanishes q times on the interval $0 \leq x < p$. The function $B(x)$ likewise vanishes q times on $0 \leq x < p$ since the zeros of $y(x)$ and $w(x)$ mutually separate each other. The series $T(y)$ has the frequency q/p as do the series $T'(y)$, $T(w)$, etc.

Let ω be a point which is not a zero of $y(x)$ nor of $w(x)$, and let $u(x)$ be a solution which vanishes at ω without being identically zero. Then $u(\omega+p) \neq 0$. Otherwise for some constant $\rho > 0$

$$u(x+p) \equiv \rho u(x).$$

As we have seen, $u(x)$ would then be a constant multiple of $y(x)$ or of $w(x)$, contrary to the hypothesis that ω is not a zero of $y(x)$ or of $w(x)$. The q -th zero ω' of $u(x)$ following ω is such that $\omega' - \omega \neq p$. If $\omega' - \omega < p$, then after a suitable change of coördinates of the form $x' = x + x_0$, the interval $0 < x' < p$ will include both ω and ω' and $T(u)$ will possess a p -chain of b -length $q+1$. If $\omega' - \omega > p$ and x is suitably chosen, the interval $0 \leq x' \leq p$ will include just $q-1$ zeros of $u(x)$, and $T(u)$ will possess a p -chain of b -length $q-1$. In either case the series $T(u)$ as well as the series $T'(u)$ is skew Sturmian, and the proof of the theorem is complete.

THE INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

- Birkhoff, G. D., *Dynamical Systems*, American Mathematical Society Colloquium Publications (1927).
 Koksma, J. F., "Diophantische Approximationen," *Ergebnisse der Mathematik und ihre Grenzgebiete*, IV, 4 (1936).
 Lévy, Paul, "Sur le développement en fraction continue d'un nombre choisi au hasard," *Compositio Mathematica*, vol. 3 (1936), pp. 286-303.
 Morse, Marston and G. A. Hedlund, "Symbolic Dynamics," *American Journal of Mathematics*, vol. 60 (1938), pp. 815-866.
 Perron, Oskar, *Die Lehre von den Kettenbrüchen*, Teubner (1929).

ON THE METHOD OF FINDING ISOTROPIC STATIC SOLUTIONS OF EINSTEIN'S FIELD EQUATIONS OF GRAVITATION.*

By P. Y. CHOU.

1. Introduction. The isotropic static solutions of Einstein's field equations can be obtained by solving a set of differential equations given by the author.¹ For the type of problem which involves the determination of isotropic fields of a single body the complete solution of the problem reduces first to the solution of a non-linear partial differential equation of the second order (I, (3.4)), and secondly to the transformation of ds^2 in the (u, v, w) coördinates to the canonical form (I, (2.4)). When the problems are simple such as the examples given in I, where the non-linear partial differential equation degenerates into an ordinary equation, these two steps can be accomplished with ease. But in the general problem the solution of the partial differential equation and the transformation of coördinates are both difficult.

An alternative way of approach is to compute the equations (2.5) in I in terms of the (x, y, z) coördinates in the canonical form (2.4). Then we have two dependent variables, U and σ , satisfying seven partial differential equations with three independent variables, x, y, z . Unfortunately these equations are non-linear in U and σ and the determination of their general solution is by no means simple.

In the present paper we shall derive from (2.5) in I, as a further necessary condition, another set of partial differential equations whose solutions also satisfy the field equations. The advantage of the present treatment over the previous one is, as we shall show presently, that we have to solve a set of seven non-linear partial differential equations of the third order satisfied by only one dependent variable σ while the other function U can be constructed out of σ and the partial derivatives of σ . Out of the seven partial differential equations we shall derive the well-known Laplace's equation. Hence as a method of procedure we can choose a harmonic function and this harmonic function will define an isotropic static gravitational field provided it satisfies the remaining six partial differential equations simultaneously. As an illustration of the present method we shall show that Kasner's solution (I, (3.5)) and the field of the semi-infinite plane (I, (3.9)) are the only two-dimensional isotropic static fields in Einstein's theory of gravitation.

* Received October 30, 1938.

¹ P. Y. Chou, *American Journal of Mathematics*, vol. 59 (1937), p. 754, which will be referred to as "I", eq. (2.5).

2. Equations determining σ . We write down some of the important equations in I. The field equations (I, (2.3)) are

$$(2.1) \quad G_{ij} = R_{ij} + U_{i,j}/U = 0, \quad G_{0i} = 0, \quad G_{00} = UU_{,i}{}^i = 0.$$

The canonical form of the isotropic static arc element (I, (2.4)) is

$$(2.2) \quad ds^2 = U^2 dt^2 - e^{-2\sigma}(dx^2 + dy^2 + dz^2).$$

The equations determining the isotropic solutions of the field equations can be written as (I, (2.5)),

$$(2.3) \quad U_{i,j} = \frac{6U}{C + U^2} (U_i U_j - \frac{1}{3} g_{ij} U^h U_h), \quad \text{and} \quad R = 0.$$

From (2.3) we can derive as an integral (I, (2.6)),

$$(2.4) \quad U^h U_h = -k^2(c + U^2)^4.$$

We can eliminate $U_{i,j}$ between (2.1) and (2.3). Then

$$(2.5) \quad R_{ij} = -\frac{6}{C + U^2} (U_i U_j - \frac{1}{3} g_{ij} U^h U_h), \quad U_{,i}{}^i = 0.$$

In fact (2.5) was obtained when we proved the sufficiency of (2.3) determining the isotropic solutions of (2.1) (I, (2.17)). Now let us form the invariant $R^{mn}R_{mn}$ from (2.5):

$$(2.6) \quad R^{mn}R_{mn} = 24(U^h U_h)^2/(c + U^2)^2.$$

By using (2.4) we obtain as a consequence

$$(2.7) \quad (c + U^2)^6 = R_{mn}R^{mn}/24k^4.$$

Equations (2.5) and (2.7) are tensor equations and, accordingly, hold in any system of coördinates. If we compute R_{ij} from the form (2.2), we find ²

$$(2.8) \quad R_{ij} = -\sigma_{,ij} + \sigma_{,i}\sigma_{,j} - g_{ij}(\Delta_2\sigma + \Delta_1\sigma)$$

where

$$\Delta_2\sigma = g^{ij}\sigma_{,ij}, \quad \Delta_1\sigma = g^{ij}\sigma_{,i}\sigma_{,j}.$$

If we calculate $\sigma_{,ij}$ explicitly, we get

$$(2.9) \quad \begin{aligned} R_{11} &= -\sigma_{xx} + \sigma_y^2 + \sigma_z^2 - \frac{1}{2}(\sigma_x^2 + \sigma_y^2 + \sigma_z^2), \\ R_{12} &= -\sigma_{xy} - \sigma_x\sigma_y, \end{aligned}$$

where we put $\sigma_{xx} = \partial^2\sigma/\partial x^2$, $\sigma_y = \partial\sigma/\partial y$, etc., and the other components of the tensor R_{ij} can be obtained by cyclic permutations of the coördinates x, y, z .

² L. P. Eisenhart, *Riemannian Geometry* (1926), p. 90, eq. (28.6).

Now we can eliminate U between (2.5) and (2.7) with the understanding that the components of R_{ij} are given by (2.9). Then we see that (2.5) are seven non-linear partial differential equations of the third order in σ . Since σ has to satisfy seven equations simultaneously, its number of solutions must be limited. But if we can obtain one solution, then the corresponding function U is determined automatically by (2.7).

The conditions (2.5) and (2.7), with R_{ij} given by (2.9), are thus necessary for the field equations (2.1) to possess isotropic solutions in empty space. They are also sufficient. For from (2.5) and (2.8) we have

$$(2.10) \quad 0 = R_{ij,k} - R_{ik,j} \\ = \frac{6}{C + U^2} (U_{i,j}U_k - U_{i,k}U_j + \frac{2}{3}g_{ij}U^hU_{h,k} - \frac{2}{3}g_{ik}U^hU_{h,j}) \\ - \frac{4U}{(C + U^2)^2} U^hU_h (g_{ij}U_k - g_{ik}U_j).$$

If we contract (2.10) by g^{ij} and U^k separately and eliminate the intermediate expression $U^hU_{h,k}$, we find (2.3) again. Then the field equations (2.1) are also satisfied by the theorem proved in I. Hence we may summarize the above results in the following theorem:

THEOREM. *A necessary and sufficient condition for the static field equations (2.1) to possess isotropic solutions in empty space is that the function σ should satisfy equations (2.5) where R_{ij} and U are given by (2.9) and (2.7) respectively.*

The theorem proved in I lays emphasis on the function U and the present theorem deals primarily with σ . The method of obtaining σ is as follows: From (2.5) and (2.9) we find

$$(2.11) \quad R = e^{2\sigma} [\sigma_{xx} + \sigma_{yy} + \sigma_{zz} - \frac{1}{2}(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)] \\ = -e^{5\sigma/2} (f_{xx} + f_{yy} + f_{zz}) = 0; \quad f = e^{-\sigma/2}.$$

In other words f satisfies the classical Laplace's equation. Since Laplace's equation possesses a large variety of solutions, the isotropic static fields are defined by those which will also satisfy the other equations in (2.5). Hence we may test whether any harmonic function f defines an isotropic static field by simply constructing R_{ij} and U according to (2.9) and (2.7) and see whether every member of the equations in (2.5) is verified.

We have dealt with the case in empty space only. A corresponding theorem within matter can also be proved in a similar way.

3. Two-dimensional problem. Although we have given the general

method of finding isotropic static fields in the previous section, it is still rather laborious to solve the general three-dimensional problem. On the other hand when we restrict ourselves to the two-dimensional case, the present method with slight modifications can give us all the isotropic fields; it is decidedly simpler than trying to solve the set of equations (2.3) directly, for even in the canonical form (2.2) when the coördinate z is absent from the functions U and σ , (2.3) are still non-linear and not much information can be obtained from these equations.

Since both σ and U are assumed to be independent of z , we have from (2.9),

$$\begin{aligned}
 R_{11} &= -\sigma_{xx} + \sigma_y^2 - \frac{1}{2}(\sigma_x^2 + \sigma_y^2) = -\frac{6}{C+U^2}[U_x^2 - \frac{1}{3}(U_x^2 + U_y^2)], \\
 R_{22} &= -\sigma_{yy} + \sigma_x^2 - \frac{1}{2}(\sigma_x^2 + \sigma_y^2) = -\frac{6}{C+U^2}[U_y^2 - \frac{1}{3}(U_x^2 + U_y^2)], \\
 R_{12} &= -\sigma_{xy} - \sigma_x\sigma_y = -\frac{6}{C+U^2}U_xU_y, \\
 R_{33} &= \frac{1}{2}(\sigma_x^2 + \sigma_y^2) = \frac{2}{C+U^2}(U_x^2 + U_y^2),
 \end{aligned}
 \tag{3.1}$$

the other components of R_{ij} all vanishing identically. According to the given procedure we should eliminate U from (3.1) by using (2.7). But this is clumsy and we shall avoid using it. Instead, from (2.4) and R_{33} in (3.1) we find

$$(3.2) \quad (c + U^2)^3 = e^{2\sigma}(\sigma_x^2 + \sigma_y^2)/4k^2.$$

Furthermore we can drop out the common expression R_{33} in R_{11} and R_{22} and get

$$\begin{aligned}
 R_{11}: \quad \sigma_{xx} + \sigma_x^2 &= 6U_x^2/(c + U^2), \\
 R_{22}: \quad \sigma_{yy} + \sigma_y^2 &= 6U_y^2/(c + U^2), \\
 R_{12}: \quad \sigma_{xy} + \sigma_x\sigma_y &= 6U_xU_y/(c + U^2).
 \end{aligned}
 \tag{3.3}$$

The harmonic function f in (2.11) can now be taken to be

$$(3.4) \quad f = e^{-\sigma/2} = F(x + iy) + F(x - iy) = F + F^*,$$

where F is an analytic function of the complex variable $x + iy$. Then the partial derivatives of σ with respect to x and y can be computed. If we denote the derivative of F with respect to its argument by \dot{F} , we find (3.2) to be

$$(3.5) \quad (c + U^2)^3 = 4\dot{F}\dot{F}^*/k^2f^6.$$

By means of (3.4) and (3.5) we can put (3.3) in the following form:

$$\begin{aligned}
 R_{11}: & \frac{2}{f} (\ddot{F} + \ddot{F}^*) - \frac{6}{f^2} (\dot{F} + \dot{F}^*)^2 = -\frac{C + U^2}{6U^2} \left[\frac{\ddot{F}}{\dot{F}} - \frac{6\dot{F}}{f} + \frac{\ddot{F}^*}{\dot{F}^*} - \frac{6\dot{F}^*}{f} \right]^2, \\
 (3.6) \quad R_{22}: & -\frac{2}{f} (\ddot{F} + \ddot{F}^*) + \frac{6}{f^2} (\dot{F} - \dot{F}^*)^2 = \frac{C + U^2}{6U^2} \left[\frac{\ddot{F}}{\dot{F}} - \frac{6\dot{F}}{f} - \frac{\ddot{F}^*}{\dot{F}^*} + \frac{6\dot{F}^*}{f} \right]^2, \\
 R_{12}: & \frac{2}{f} (\ddot{F} - \ddot{F}^*) - \frac{6}{f^2} (\dot{F}^2 - \dot{F}^{*2}) = -\frac{C + U^2}{6U^2} \left[\left(\frac{\ddot{F}}{\dot{F}} - \frac{6\dot{F}}{f} \right)^2 - \left(\frac{\ddot{F}^*}{\dot{F}^*} - \frac{6\dot{F}^*}{f} \right)^2 \right].
 \end{aligned}$$

Then we form linear combinations of the above equations. The expression $R_{11} + R_{22}$ gives

$$(3.7) \quad \frac{36}{f^2} \dot{F}\dot{F}^* = \frac{C + U^2}{U^2} \left(\frac{\ddot{F}}{\dot{F}} - \frac{6\dot{F}}{f} \right) \left(\frac{\ddot{F}^*}{\dot{F}^*} - \frac{6\dot{F}^*}{f} \right)$$

which in combination with (3.5) also gives (2.4). From $R_{11} - R_{22}$ we find

$$\begin{aligned}
 (3.8) \quad \frac{2}{f} (\ddot{F} + \ddot{F}^*) - \frac{6}{f^2} (\dot{F}^2 + \dot{F}^{*2}) \\
 = -\frac{C + U^2}{6U^2} \left[\left(\frac{\ddot{F}}{\dot{F}} - \frac{6\dot{F}}{f} \right)^2 + \left(\frac{\ddot{F}^*}{\dot{F}^*} - \frac{6\dot{F}^*}{f} \right)^2 \right].
 \end{aligned}$$

From R_{12} in (3.6) and (3.8) we can easily see that the following equation and its conjugate must hold:

$$(3.9) \quad \frac{2}{f} \ddot{F} - \frac{6}{f^2} \dot{F}^2 = -\frac{C + U^2}{6U^2} \left(\frac{\ddot{F}}{\dot{F}} - \frac{6\dot{F}}{f} \right)^2$$

which can be simplified into

$$U^2 f^2 \ddot{F}^2 / \dot{F}^4 + c(f\ddot{F}/\dot{F}^2 - 6)^2 = 0.$$

Since only U^2 presents itself in the arc element (2.2), we may take the positive sign in front of U , namely,

$$(3.10) \quad f\ddot{F}/\dot{F}^2 = 6\sqrt{-c}/(U + \sqrt{-c}).$$

Taking (3.9) and its conjugate, we can eliminate U^2 from (3.7) and obtain

$$\begin{aligned}
 (3.11) \quad \left(\frac{f\ddot{F}}{\dot{F}^2} - 3 \right) \left(\frac{f\ddot{F}^*}{\dot{F}^{*2}} - 3 \right) = 9, \text{ or} \\
 f \frac{d}{dz} \left(\frac{1}{\dot{F}} \right) \frac{d}{dz^*} \left(\frac{1}{\dot{F}^*} \right) + 3 \left(\frac{d}{dz} \frac{1}{\dot{F}} + \frac{d}{dz^*} \frac{1}{\dot{F}^*} \right) = 0.
 \end{aligned}$$

Inserting \ddot{F} from (3.10) into (3.11) and simplifying, we find finally

$$(3.12) \quad [\sqrt{-c} + (\sqrt{-c})^*] U = 0.$$

In other words, the arbitrary constant c must either be zero or a positive constant which can be taken to be unity without the loss of generality.

Case (1). $c = 0$. Then from (3.10), we find

$$(3.13) \quad \ddot{F} = 0, \quad F = bz/2, \quad f = bx \text{ and } e^{-2\sigma} = b^4 x^4.$$

From (3.5) we get $U^2 = 1/b^2 x^2$ so that $k = b$. This is Kasner's solution for an infinite plane (I, (3.5)).

Case (2). $c = 1$. Then F must be different from zero and (3.11) can be written as:

$$(3.14) \quad F + F^* + 3 \left(\frac{dz}{dF} / \frac{d^2 z}{dF^2} + \frac{dz^*}{dF^*} / \frac{d^2 z^*}{dF^{*2}} \right) = 0.$$

Since z and z^* are actually independent, we may conclude that:

$$(3.15) \quad F + 3 \frac{dz}{dF} / \frac{d^2 z}{dF^2} = i\alpha,$$

where α must be a real constant. In fact α can be put equal to zero without any loss of generality, for it only adds an imaginary constant to F and in the final expression of f in (3.4), it does not appear. Hence integrating (3.15), we find

$$(3.16) \quad F = -(\beta z)^{-1/2},$$

in which β is a constant of integration.

We have two expressions of U from (3.10) and (3.5). They must be identical so that $\beta\beta^* = 64k^2$. Then

$$(3.17) \quad U^2 = \tan^2(\phi + \delta)/2, \quad e^{2\sigma} = f^4 = \frac{1}{(2k)^2 \rho^2} \cos^4(\phi + \delta)/2,$$

where δ is the amplitude of the complex number β , ϕ and ρ are the amplitude and modulus of $x + iy$. This represents the field of the semi-infinite plane (I, (3.9)). In other words Kasner's solution and the field of the semi-infinite plane are the only two-dimensional isotropic static fields in empty space according to Einstein's theory of gravitation.

THE NATIONAL SOUTH-WEST ASSOCIATED UNIVERSITY (being a wartime union of National Tsing Hua University and National Peking University of Peiping, and Nankai University of Tientsin),

KUNMING, YÜNNAN PROVINCE, CHINA.

ON THE ALMOST PERIODIC BEHAVIOR OF THE LUNAR NODE.*

By AUREL WINTNER.

Introduction. In view of Newton's deduction of his approximation formula for the mean motion, ω , of the ascending node of the lunar path (cf., e. g., Tisserand [6], pp. 42-44), the constant ω may be characterized, from the astronomical point of view, not only as the average velocity of the nodal angle $\vartheta = \vartheta(t)$, but also in terms of the relative number of times the Moon passes through the ecliptic (on the average). Correspondingly, the precise form of Newton's approximation theory, as developed by Adams by means of infinite determinants, is directly based on the Jacobian differential equation which determines the ordinate $z = z(t)$, if the ecliptic is the (x, y) -plane (cf., e. g., Tisserand [6], pp. 286-288).

From the mathematical point of view, there immediately arise several questions. Some of these have been investigated by Levi-Civita [4], who proved that the two definitions of ω (those based on $\vartheta(t)$ and $z(t)$, respectively) are equivalent, and that the limit which ω is supposed to represent actually exists; so that, in the theory of Adams,

$$(1) \quad \vartheta(t) = \omega t + \psi(t), \text{ where } \omega = \text{const.} \neq 0 \text{ and } |\psi(t)| < \text{Const.}$$

The present paper deals with certain analytical refinements of Levi-Civita's result; refinements which, though of apparent astronomical significance, can only be treated by using analytical tools developed recently (Levi-Civita's paper appeared in 1911).

The modern theory of the Moon, as originated by Hill and further developed by Brown (cf., e. g., Poincaré [5]), is based on certain tacit assumptions which, for the case at hand, imply that the non-secular part of $\vartheta(t)$, i. e., the remainder term $\psi(t)$ of (1), may be analyzed into an anharmonic Fourier series. This assumption will be justified by proving that $\psi(t)$ is almost periodic (almost periodicity will always be meant in the sense of Bohr). The formal situation is as follows:

The variational equation of Adams for the ordinate z is of the form

$$(2) \quad z'' + f(t)z = 0,$$

where $f(t)$ is a given periodic function of the time. One can write this differential equation of the second order in the form

* Received January 3, 1939.

$$(3) \quad u' = a(t)u + b(t)v, \quad v' = c(t)u + d(t)v$$

of two differential equations of the first order. Then, on introducing into the (u, v) -plane polar coördinates by placing

$$(4) \quad u = (u^2 + v^2)^{\frac{1}{2}} \cos \vartheta, \quad v = (u^2 + v^2)^{\frac{1}{2}} \sin \vartheta,$$

one can conclude from a general theorem, that not only is $\vartheta(t)$ of the form (1) but, in addition, $\psi(t)$ is almost periodic (cf. Wintner [9]).

However, this approach to the problem is based on the assumption that the coördinate $\vartheta = \vartheta(t)$ of the ascending node of the Moon is identical with the angle $\vartheta = \vartheta(t)$ which is defined by (4). Now, the latter ϑ , being the polar angle in the (u, v) -plane, clearly is not the lunar node, if one writes (2) in the form (3) by placing $u = z, v = z'$ (or $u = z', v = z$). Fortunately, it turns out that the polar angle in the (u, v) -plane becomes identical with the nodal coördinate $\vartheta = \vartheta(t)$ of the Moon if, instead of identifying u, v with z, z' , one subjects the pair z, z' to a suitable linear substitution whose matrix is a certain periodic function of t , and then defines u, v as the resulting linear combinations of z, z' .

Due to the relations which Levi-Civita used when identifying the two definitions of ω (cf. the beginning of this paper), the linear transformation defining u, v is quite explicit. Correspondingly, the existence of ω and the almost periodicity of $\psi(t)$ will be proved directly. This direct proof will not involve an actual modification of the program sketched above. In fact, the proof depends, in either case, on an application of the following theorem, formulated as a conjecture by the present author, and subsequently proved by Bohr [1]:

If $\vartheta(t)$ is real and $i\vartheta(t)$ almost periodic, then there exist a constant ω and an almost periodic function $\psi(t)$ such that $\vartheta(t) = \omega t + \psi(t)$. (The converse of this theorem is obvious.)

Since the proofs will be based on the theory of almost periodic functions, the proofs are independent of the results of Levi-Civita (cf. the beginning of this paper), which, therefore, follow as corollaries.

The almost periodicity of the remainder term $\psi(t)$ will also imply the existence of an asymptotic distribution function for the angular variable $\vartheta(t)$. This means that there exists an asymptotic probability $p = p(\alpha, \beta)$, that the lunar node $\vartheta(t)$ will lie on a given arc $\alpha \leq \vartheta(t) \leq \beta$, where $\vartheta = \vartheta(t)$ is thought of as reduced mod 2π . Needless to say, (1) in itself would be insufficient to guarantee the existence of such an asymptotic distribution function.

1. The considerations of this section are more general than those actually

needed for the problem at hand, and concern the explicit characterization of all (non-conservative) linear canonical transformations in case of n degrees of freedom. The result, though of algebraic simplicity, does not seem to occur in the classical literature of the subject, apparently because the verification requires some effort, if one starts out with the standard Pfaffian criterion of canonical transformations. On the other hand, the result follows quite naturally by using a method developed recently (cf. Wintner [8], van Kampen and Wintner [7]).

Reserving the sign ' for d/dt , let A^* denote the transposed matrix of the matrix A (although all matrices occurring will be real); so that $A'^* = A^*$, where $A = A(t)$. Correspondingly, the bilinear form belonging to a matrix B will be denoted by Y^*BX , where X, Y are real column vectors. Let E be the n -rowed unit matrix, O the n -rowed zero matrix, and I the $2n$ -rowed skew-symmetric matrix

$$(5) \quad I = \begin{pmatrix} O & E \\ -E & O \end{pmatrix}; \text{ so that } I^{-1} = I^* = -I, \quad \det I = +1.$$

Then the most general linear (homogeneous) Hamiltonian system with n degrees of freedom is

$$(6) \quad IX' = S(t)X,$$

where $S(t)$ is an arbitrarily given, $2n$ -rowed, symmetric, possibly singular, continuous matrix function of the time t . In fact, if x_1, \dots, x_{2n} denote the components of the vector X , one can write (6) as

$$x'_i = -\partial H / \partial x_{i+n}, \quad x'_{i+n} = \partial H / \partial x_i, \quad (i = 1, \dots, n),$$

where $H = H(X; t)$ is the quadratic form $H = \frac{1}{2}X^*S(t)X$; so that x_{i+n} , where $i = 1, \dots, n$, is the i -th coördinate, and x_i the momentum canonically conjugate to x_{i+n} .

If one subjects X to an arbitrary linear substitution

$$(7) \quad \bar{X} = T(t)X,$$

where $T(t)$ is a matrix function of $(2n)^2$ elements which have continuous first derivatives and a non-vanishing determinant, then (6) clearly is transformed into a system of $2n$ differential equations of the first order which are again homogeneous and linear and can, therefore, be written in the form

$$(8) \quad I\bar{X}' = \bar{S}(t)\bar{X},$$

the matrix (5) being non-singular; so that $\bar{S}(t)$ is uniquely determined by $S(t)$, $T(t)$ and the derivative matrix $T'(t)$. However, the transform (8) of

the canonical system (6), is not, in general, again canonical, since $\bar{S}(t)$ need not be symmetric matrix whenever $S(t)$ is symmetric. Correspondingly, (7) may be defined to be a canonical transformation if it has the property that (8) is a canonical system for every canonical system (6), i. e., if $\bar{S}^*(t) = \bar{S}(t)$ whenever $S^*(t) = S(t)$.

Now, $T(t)$ will have this property if and only if there exists a scalar constant $\mu \neq 0$ such that

$$(9) \quad T^*(t)IT(t) = \mu I \quad \text{for every } t \quad (\mu' \equiv 0).$$

(It may be mentioned that (9) implies that

$$(9 \text{ bis}) \quad \det T(t) = \mu^n, \quad (\det T(t) \neq 0),$$

and not only that $|\det T(t)| = |\mu|^n$). Furthermore, if the necessary and sufficient condition (9)–(9 bis) for a canonical linear transformation (7) is satisfied, the matrix $\bar{S}(t)$ of the transformed Hamiltonian function, i. e., of the quadratic form $\frac{1}{2}\bar{X}^*\bar{S}(t)\bar{X}$, follows from

$$(10) \quad T^*\bar{S}T = \mu S + T^*IT', \quad \text{where } S = S(t), T = T(t), \det T(t) \neq 0;$$

(so that (10) is a symmetric matrix for every symmetric $S(t)$ if, and only if, (9) is satisfied).

The proof of the statements (9), (10) will be omitted. For, on the one hand, these statements may be verified from the general (non-linear) results, obtained loc. cit. [7], at least if one disregards the fact that, this time, only the existence of a first continuous derivative $T'(t)$ is required (the problem being linear). And, on the other hand, a direct verification proceeds in exactly the same way as in the particular case $T(t) = \text{const}$, treated loc. cit. [8].

2. Suppose, in particular, that (7) transforms momenta into momenta and coördinates into coördinates; so that

$$(11) \quad T(t) = \begin{pmatrix} A(t) & O \\ O & B(t) \end{pmatrix},$$

where $A(t)$, $B(t)$ are non-singular n -rowed matrices. It is easily verified from (5) that (9) is satisfied by (11) and $\mu = +1$ if and only if $A^*(t) = B^{-1}(t)$.

It follows that in the particular case $A = B$ of a cogredient transformation of the momenta and coördinates, the condition is that $A(t)$ be, for every t , an n -rowed orthogonal matrix (of determinant ± 1). Application of (10) to this particular case shows that the transformed Hamiltonian function is

$$(12) \quad \frac{1}{2}\bar{X}^*\bar{S}(t)\bar{X} = \frac{1}{2}X^*S(t)X + \frac{1}{2}\{U^*A'(t)A^*(t)V - V^*A'(t)A^*(t)U\},$$

where U and V are the vectors with n components, (\bar{x}_i) and (\bar{x}_{i+n}) , which are formed by the n first and n last components of the vector (7) with $2n$ components.

For instance, if the degree of freedom is an even number, say $n = 2m$, condition (8) is satisfied (with $\mu = +1$) if $T(t)$ is the $2n$ -rowed matrix which one obtains by repeating, $2m$ times along the principal diagonal, any two-rowed rotation matrix

$$(13) \quad \Phi(t) = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}, \text{ where } \phi = \phi(t)$$

is any given scalar function which has a continuous derivative $\phi'(t)$. In this case, the quadratic form $\frac{1}{2} \{ \}$, which in (12) represents the deviation of the new and old Hamiltonian functions, readily reduces to

$$(14) \quad \frac{1}{2} \bar{X}^* \bar{S}(t) \bar{X} - \frac{1}{2} X^* S(t) X = \phi'(t) \sum_{k=1}^m (\xi_k H_k - \eta_k \Xi_k),$$

if $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \dots, \bar{x}_{2n-1}, \bar{x}_{2n}$ are respectively denoted by $\Xi_1, H_1, \Xi_2, H_2, \dots, \xi_m, \eta_m$, where $m = \frac{1}{2}n$.

3. Let the degree of freedom be $n = 1$, and write $p, q; u, v$ for $x_1, x_2; \bar{x}_1, \bar{x}_2$ respectively; so that (7) becomes

$$(15) \quad \begin{aligned} u &= \alpha(t)p + \beta(t)q, \\ v &= \gamma(t)p + \delta(t)q, \end{aligned} \quad \text{where} \quad \begin{pmatrix} \alpha(t) & \beta(t) \\ \gamma(t) & \delta(t) \end{pmatrix} = T(t).$$

It is easily verified from (5), where $E = 1, O = 0$ in the present case, that the necessary and sufficient condition (9) for a linear canonical transformation is satisfied by (15) if and only if $\det T(t) = \text{const.}$ But $\text{const.} = \mu$, by (9 bis); so that the criterion takes the form

$$(16) \quad \alpha(t)\delta(t) - \beta(t)\gamma(t) = \mu, \text{ where } \mu = \text{const.} \neq 0.$$

It follows, therefore, by straightforward reductions that

$$(17) \quad 4\mu IT'(t)T^{-1}(t) = \begin{pmatrix} 2\Delta_{\gamma\delta}(t) & \Delta_{\beta\gamma}(t) - \Delta_{\alpha\delta}(t) \\ \Delta_{\beta\gamma}(t) - \Delta_{\alpha\delta}(t) & 2\Delta_{\alpha\beta}(t) \end{pmatrix},$$

if the elements of this two-rowed matrix denote the determinants defined by

$$(18) \quad \Delta_{\kappa\lambda}(t) = \kappa'(t)\lambda(t) - \lambda'(t)\kappa(t), \text{ where } v' = dv/dt.$$

If, in particular, the constant (16) is 1, then, on denoting the new and old Hamiltonian functions, i. e., the quadratic forms $\frac{1}{2} \bar{X}^* \bar{S}(t) \bar{X}$ and $\frac{1}{2} X^* S(t) X$, by $K(u, v; t)$ and $H(p, q; t)$, one sees from (10), (7) and (17) that

$$(19) \quad K(u, v; t) = H(p, q; t) + \frac{1}{2}\{\Delta_{\gamma\delta}u^2 + (\Delta_{\beta\gamma} - \Delta_{\alpha\delta})uv + \Delta_{\alpha\beta}v^2\},$$

where $H(p, q; t)$ is thought of as expressed by means of the inverse of the substitution (15) as a function of $(u, v; t)$, and the Δ are, in view of (18), given functions of t .

4. Let

$$(20) \quad x'' - 2y' = \Omega_x(x, y, z), \quad y'' + 2x' = \Omega_y(x, y, z), \quad z'' = \Omega_z(x, y, z)$$

be the equations of motion of the (non-planar) restricted problem of three bodies in a synodical barycentric coördinate system (x, y, z) ; so that the axis of syzygies is the x -axis which rotates, with reference to a sidereal planar coördinate system which coincides with the (x, y) -plane, with constant angular velocity. Thus, (20) is an irreversible, conservative, dynamical system with three degrees of freedom, admitting Jacobi's integral of relative energy:

$$(21) \quad \frac{1}{2}(x'^2 + y'^2 + z'^2) - \Omega(x, y, z) = \text{const.}$$

The classical mathematical literature of the restricted problem of three bodies concerns the case $z(t) \equiv 0$ of a planar solution

$$(22) \quad x = x(t), \quad y = y(t).$$

Starting with any given planar solution (22), consider the non-planar solutions in the infinitesimal neighborhood of (22); so that the third of the equations (20) may be replaced by its Jacobi equation belonging to (22). Then $z = z(t)$ is determined by the equation (2) of Adams, in which the coefficient function $f(t)$ is obviously given by

$$(23) \quad f(t) = -\Omega_{zz}(x(t), y(t), 0).$$

Furthermore, on denoting by $\vartheta = \vartheta(t)$ the longitude of the ascending node, and by $\iota = \iota(t)$ the (small) inclination, with reference to the synodical coördinate system (x, y) , one has in (2)

$$(24) \quad z = -x \sin \iota \sin \vartheta + y \sin \iota \cos \vartheta, \quad z' = -x' \sin \iota \sin \vartheta + y' \sin \iota \cos \vartheta,$$

at least so long as

$$(25) \quad x(t)y'(t) - y(t)x'(t) \neq 0.$$

In order to see this, it is sufficient to write down, within the degree of accuracy of (2), the projections of the vector product of (x, y, z) and (x', y', z') on the coördinate axes; cf. Levi-Civita [4], pp. 366-367 (where, however, the ascending node is referred to the sidereal, instead of the synodical, coördinate system; so that one has to replace $\vartheta(t)$ by $\vartheta(t) - t$).

5. In view of (24), where x, y are the given functions (22) of t , one can replace the differential equation (2) of the second order for the ordinate z by two differential equations of the first order for the Eulerian angles ι, ϑ . In order to legalize this, it is sufficient to observe that, barring the case of the trivial solution $z(t) \equiv 0$ of (2) which belongs to the given generating solution (22) of (20), the angles $\iota = \iota(t)$, $\vartheta = \vartheta(t)$ do not become undetermined, since

$$(26) \quad \sin \iota \neq 0$$

for every t . For if (26) were violated at some $t = t_0$, it would follow from (24) that $z = 0$ and $z' = 0$ at this particular t . But then (2) implies that $z = 0$ for every t .

6. It turns out that the differential equations for ι, ϑ , mentioned at the beginning of § 5, readily lead to the desired equations (3) in which u, v represent certain linear combinations of z, z' in such a way that the requirement (4) of the Introduction becomes satisfied.

To this end, put

$$(27) \quad u = (xy' - yx')^{\frac{1}{2}} \sin \iota \cos \vartheta, \quad v = (xy' - yx')^{\frac{1}{2}} \sin \iota \sin \vartheta,$$

if the determinant (25) is positive, and modify the factors $(xy' - yx')^{\frac{1}{2}}$ on the left of (27) in an obvious manner, if the continuous non-vanishing function (25) of t is negative. According to (27), one can write (24) in the form

$$(28) \quad z = (xy' - yx')^{\frac{1}{2}}(yu - xv), \quad z' = (xy' - yx')^{\frac{1}{2}}(y'u - x'v)$$

of a linear substitution of u, v into z, z' . The coefficient matrix of this linear substitution is, by (22), a known function of t and has, in view of (28), the determinant $+1$ for every t . Hence, on placing $p = z$, $q = z'$, and writing (28) in the form (15), the condition (16) for a canonical transformation is satisfied by $\mu = 1$. On the other hand, (2) may be written in the form

$$(29) \quad p' = -\partial H / \partial q, \quad q' = \partial H / \partial p,$$

if one puts

$$(30) \quad H \equiv H(p, q; t) = -\frac{1}{2}q^2 - \frac{1}{2}f(t)p^2, \text{ where } p = z, q = z'.$$

Consequently, the representation of (2) in terms of the variables (27) is the linear canonical system

$$(31) \quad u' = -\partial K / \partial v, \quad v' = \partial K / \partial u,$$

where the Hamiltonian function $K = K(u, v; t)$ is a quadratic form in (u, v)

and is explicitly given by (19) and (30), the determinants (18) being obtained by identifying (15) with (28).

Finally, on identifying (31) with (3), one sees from (27) that the requirement (4) of the Introduction is satisfied, and one has

$$(32) \quad u^2 + v^2 = (xy' - yx') \sin^2 t.$$

It should be mentioned for later application that the pair of conditions (25), (26) is, in view of (32), equivalent to the condition that $u = u(t)$, $v = v(t)$ do not vanish simultaneously; a condition which, in turn, is equivalent to the exclusion of the trivial solution $u(t) \equiv 0$, $v(t) \equiv 0$ of (3), i. e., of (31).

7. Without assuming that the (real) system (3) has the canonical form (31), suppose that its coefficient functions $a(t), \dots, d(t)$ have a common period, say τ . Suppose further that the characteristic exponents of (3) are of the non-degenerate stable type, i. e., that the pair of the characteristic roots of the monodromy group is of the form $(\rho, 1/\rho)$, where $|\rho| = 1$ but $\rho \neq \pm 1$. Then it is readily seen from the Fuchs-Floquet representation of the general solution of (3), that (3) admits a fundamental matrix which is the product of two real matrices of the following type: One of these two matrix functions of t is periodic, with τ as period, while the other matrix factor not only is periodic but represents a uniform rotation, with a period which is determined by the characteristic exponent, i. e., by $\arg \rho$.

Now, the existence of a fundamental matrix which possesses a factorization of this type clearly implies, not only that every solution $u = u(t)$, $v = v(t)$ of (3) is almost periodic, but also that the greatest lower bound of $u^2 + v^2$ for $-\infty < t < +\infty$ is distinct from zero for all those (real) solutions $u = u(t)$, $v = v(t)$ of (3) for which $u^2 + v^2 = 0$ does not hold at some fixed $t = t_0$.

Consequently, $u = u(t)$ and $v = v(t)$ cannot simultaneously come arbitrarily close to 0 for $-\infty < t < +\infty$, if one excludes the trivial solution $u(t) \equiv 0$, $v(t) \equiv 0$.

8. Now consider the case in which the given planar solution (22) of (20) is periodic. Then so is the coefficient function (23) of (2) and, therefore, the coefficient matrix of (29), or of (31). If, in particular, (22) is that solution of the restricted problem of three bodies which corresponds to Hill's intermediary lunar orbit of his limiting case, then (25) is known to be satisfied, and the characteristic exponent of (2), i. e., of (31), fulfils the stability condition required at the beginning of § 7 (as to the numerical situation, cf. Tisserand [6], p. 288).

It follows, therefore, from § 7 that

(i) the solutions $u = u(t)$, $v = v(t)$ of (31) are almost periodic and have frequencies contained in the integral modul of two numbers, say of λ and ν , where it is understood that λ and ν are or are not linearly dependent according as the period of (22) does or does not satisfy a commensurability condition with reference to the characteristic exponent;

(ii) barring the trivial solution $u(t) \equiv 0$, $v(t) \equiv 0$, one has

$$(u(t))^2 + (v(t))^2 > \text{const.} > 0 \text{ for } -\infty < t < +\infty,$$

where the const. depends on the integration constants of the solution $u = u(t)$, $v = v(t)$ of (31).

On comparing (ii) with (27), (32), and using from (i) only the fact that $u = u(t)$ and $v = v(t)$ are almost periodic, one sees that $\exp i\vartheta(t)$ is almost periodic. It follows, therefore, from the general theorem of Bohr, mentioned in the Introduction, that (1) holds for a certain constant ω and for a certain almost periodic function $\psi(t)$.

If, in addition, use is made of the description (i) of the moduli of $u(t)$ and $v(t)$, it also follows that ω and the frequencies of $\psi(t)$ are contained in the integral modul generated by the pair of numbers λ, ν (which may be commensurable); cf. Bohr [2].

The actual values of the integers j, k for which $j\lambda + k\nu$ becomes the mean motion ω readily follow, for mere reasons of continuity, from an inspection of Newton's approximation, i. e., of the problems of two bodies (cf. Levi-Civita [4], p. 376).

9. In view of the significance of the lunar node, it is natural to ask, how are the values of the angle $\vartheta(t)$ distributed asymptotically along the boundary of a circle $0 < \theta \leq 2\pi$. More precisely, the question concerns the existence (and then the determination) of a function $\sigma = \sigma(\theta)$, the angular asymptotic distribution function, which is defined for $0 < \theta \leq 2\pi$ as follows: If $L_T(\theta)$ denotes the sum of the lengths of those t -intervals which, on the one hand, are contained in the range $0 \leq t \leq T$ and, on the other hand, are such that on their points t the (continuous) angular function $\vartheta(t)$, when reduced mod 2π , satisfies the inequalities $0 < \vartheta(t) \leq \theta$, then there exists on $0 < \theta \leq 2\pi$ a monotone function $\sigma(\theta)$ which satisfies the relation $\sigma(2\pi) - \sigma(+0) = 1$ and is such that relative amount of time represented by the ratio $L_T(\theta) : T$ tends, as $T \rightarrow +\infty$, to the limit $\sigma(\theta)$ at every continuity point θ of σ .

It is known (cf. Haviland [3]) that a given $\vartheta = \vartheta(t)$ has an angular asymptotic distribution function $\sigma = \sigma(\theta)$ if and only if all the time averages

$$(33) \quad M\{\exp in\vartheta(t)\}, \quad \text{where } n = 0, 1, 2, \dots$$

and $M\{g(t)\} = \lim_{T \rightarrow \infty} \int_0^T g(t) dt / T$, exist, in which case $\sigma(\theta)$ may be determined as the solution of the trigonometric momentum problem,

$$(34) \quad \int_0^{2\pi} e^{in\theta} d\sigma(\theta) = M\{\exp in\vartheta(t)\}; \quad (n = 0, 1, 2, \dots).$$

Now, since $\exp i\vartheta(t)$, hence also $\exp in\vartheta(t)$, is almost periodic, the time averages (33) exist, as does, therefore, $\sigma(\theta)$.

10. In view of (i), § 8, the actual determination of σ is or is not an "elementary" task according as λ and ν are or are not commensurable.

In the first case, $\exp i\vartheta(t)$ is a periodic function; so that the asymptotic averages (33) reduce to averages over a finite t -range, and so $\sigma(\theta)$ simply follows from (34) by the inversion process which expresses the Lebesgue integrals as Stieltjes integrals.

In the second case, the content of (i), § (8), may be expressed, with a suitable choice of notation, as follows: If Θ is the torus $0 < \phi_1 \leq 1, 0 < \phi_2 \leq 1$ which is obtained from a Euclidean (ϕ_1, ϕ_2) -plane by reduction mod 1, then there exists on Θ a continuous function of the position, say $F = F(\phi_1, \phi_2)$, in such a way that either $\vartheta(t) = F(\omega t, t)$ or

$$(35) \quad \vartheta(t) = \omega t + F(\omega t, t),$$

where ω is an irrational number.

It will be sufficient to consider the case (35). Then, by Weyl's corollary to Kronecker's approximation theorem, the time average (33) may be expressed as the space average of $\exp in\phi_1 + F(\phi_1, \phi_2)$ over the torus Θ . Hence, (34) becomes

$$(36) \quad \int_0^{2\pi} \exp in\theta d\sigma(\theta) = \int_0^1 \int_0^1 \exp in\{\phi_1 + F(\phi_1, \phi_2)\} d\phi_1 d\phi_2; \quad (n = 0, 1, \dots).$$

Now, on writing the Lebesgue double integral on the left of (36) as a Stieltjes

simple integral, one sees from the uniqueness theorem of the trigonometric momentum problem that the angular asymptotic distribution function $\sigma(\theta)$ is the area of the set of those points (ϕ_1, ϕ_2) of Θ on which the function $\phi_1 + F(\phi_1, \phi_2)$, when reduced mod 2π so as to lie between 0 and 2π , attains values which do not exceed θ .

11. It is seen by comparison of the cases mentioned at the beginning of § 10, that the description of the angular asymptotic distribution function of an almost periodic function $\exp i\vartheta(t)$ is a problem of Diophantine intricacy. This situation is strikingly illustrated by the following consideration (which, however, cannot be applied to the problem (35) at hand).

Let

$$(37) \quad \psi(t) = \sum_m a_m \cos (\lambda_m t - \alpha_m)$$

be any real almost periodic function, and ω any real number which is not a linear combination (with integral coefficients) of the frequencies λ_m of $\psi(t)$. Then the angular asymptotic distribution of $\vartheta(t) = \omega t + \psi(t)$ is the equidistribution; so that $\sigma(\vartheta)$ is the linear functions $\theta : 2\pi$, no matter what is the remainder term (37) of $\vartheta(t) - \omega t$.

In order to prove this, it is, in view of (34), sufficient to show that

$$0 = M\{\exp in\vartheta(t)\} \quad \text{for } n = 1, 2, \dots,$$

since $\frac{1}{2\pi} \int_0^{2\pi} e^{in\theta} d\theta = 0$ for $n = 1, 2, \dots$. But $M\{\exp in\omega t\} = 0$ for

$n = 1, 2, \dots$; while $\vartheta(t) = \omega t + \psi(t)$, where $\psi(t)$ is given by (37).

Consequently, it is sufficient to show that

$$\begin{aligned} M\{\exp (in\omega t)\} M\{\exp (in \sum_m a_m \cos (\lambda_m t - \alpha_m))\} \\ = M\{\exp in(\omega t + \sum_m a_m \cos \lambda_m(t - \alpha_m))\}. \end{aligned}$$

Now, the truth of the last relation may readily be verified, for every n , from the assumption that ω is linearly independent of the λ_m .

REFERENCES

1. H. Bohr, "Kleinere Beiträge zur Theorie der fastperiodischen Funktionen, I", *Det Kgl. Danske Videnskabernes Selskab Math.-Fys. Meddelelser*, vol. 10, no. 10 (1930).
2. H. Bohr, "Ueber fastperiodische ebene Bewegungen," *Commentarii Mathematici Helvetici*, vol. 4 (1932), pp. 51-64.
3. E. K. Haviland, "On statistical methods in the theory of almost-periodic functions," *Proceedings of the National Academy of Sciences*, vol. 19 (1933), pp. 549-555.
4. T. Levi-Civita, "Sur les équations linéaires à coefficients périodiques et sur le moyen mouvement du noeud lunaire," *Annales de l'Ecole Normale Supérieure*, ser. 3, vol. 28 (1911), pp. 325-376. Cf. also Libera Trevisani (Mrs. Levi-Civita), "Sul moto medio dei nodi nel problema dei tre corpi," *Atti del Reale Istituto Veneto di scienze, letteri ed arti*, vol. 71₂ (1911-1912), pp. 1089-1137.
5. H. Poincaré, "Sur les équations du mouvement de la lune," *Bulletin Astronomique*, vol. 17 (1900), pp. 167-204.
6. F. Tisserand, *Traité de Mécanique Céleste*, vol. III, Paris (1894).
7. E. R. van Kampen and A. Wintner, "On the canonical transformations of Hamiltonian systems," *American Journal of Mathematics*, vol. 58 (1936), pp. 851-863.
8. A. Wintner, "On the linear conservative dynamical systems," *Annali di Matematica*, ser. 4, vol. 13 (1934), pp. 105-112.
9. A. Wintner, "Ueber eine Anwendung der Theorie der fastperiodischen Funktionen auf das Levi-Civitasche Problem der mittleren Bewegung," *ibid.*, vol. 10 (1931-1932), pp. 277-282.

REMARKS ON A CONJECTURE OF MINKOWSKI.*

By D. DERRY.

Let

$$L_1(x) = a_{11}x_1 + \cdots + a_{1n}x_n$$

.

.

$$L_n(x) = a_{n1}x_1 + \cdots + a_{nn}x_n$$

be n linear forms with rational coefficients and determinant ± 1 . A well known conjecture of Minkowski states that if no integral valued solution x_1, x_2, \cdots, x_n exists, other than the solution in which all the x 's are zero, for which $|L_1(x)| < 1, \cdots, |L_n(x)| < 1$ then the forms after a possible unimodular transformation of the x 's and a rearrangement of order have the form

$$L_1(x) = x_1$$

$$L_2(x) = a_{21}x_1 + x_2$$

.

.

$$L_n(x) = a_{n1}x_1 + \cdots + x_n.$$

Mordell has shown¹ that the conjecture may be stated in another form. Let $L_1(x), L_2(x), \cdots, L_n(x)$ be linear forms with rational coefficients and unit determinant which satisfy

Condition 1. For any set of integral values x_1, x_2, \cdots, x_n , other than the set in which all the x 's are zero, at least one of the forms takes a non-zero integral value.

By Mordell's result the Conjecture assumes

Form 1. At least one of the forms has integral coefficients with no common factor.

Let p be a prime which is henceforth fixed. We assume the forms also satisfy

* Received April 7, 1939.

¹ L. J. Mordell, "Minkowski's theorems and hypotheses on linear forms," Oslo Congress 1936. Form 1, communicated to me verbally by Mr. Davenport, differs slightly from the form given by Mordell. Mordell states the hypothesis in terms of the reciprocal matrix of the forms and the conjecture itself in terms of the original forms. This form states both the conjecture and the hypothesis in terms of the reciprocal matrix.

Condition 2. The coefficients of the forms are rational numbers whose denominators are some power of the prime p .

This note considers further equivalent forms of the Conjecture when Condition 2 is satisfied. These are stated in terms of finite Abelian groups and their normal series.

THEOREM 1. *Without restriction in generality the forms $L_r(x)$ used in Form 1 may be taken to have the form*

$$\begin{aligned} L_1(x) &= p^{r_1}x_1 \\ L_2(x) &= a_{21}x_1 + p^{r_2}x_2 \\ &\vdots \\ L_n(x) &= a_{n1}x_1 + \cdots + p^{r_n}x_n \end{aligned}$$

where r_1, r_2, \dots, r_n are integers with $r_1 \leq r_2 \leq \dots \leq r_n$, $p^{r_1}p^{r_2}\cdots p^{r_n} = 1$ and $p^{-r_j}a_{jk}$ integral for $j, k \geq 1$.

Proof. The above special form is easily derived² from any system of forms with unit determinant satisfying Condition 2 by rearranging the order of the forms and subjecting the x 's to unimodular transformations.

THEOREM 2. $L_1(x), L_2(x), \dots, L_n(x)$ are a set of forms with unit determinant satisfying Conditions 1 and 2 and with the form of Theorem 1. Then the values the forms assume for a set of integral values x_1, x_2, \dots, x_n are either all multiples of p^{r_n} or at least one of the forms assumes an integral value which is not a multiple of p^{r_n} .

Proof. For a set of integral values x_1, x_2, \dots, x_n , let L_1, L_2, \dots, L_n be the values taken by the forms $L_1(x), L_2(x), \dots, L_n(x)$ respectively. Let L_k be the value with the least subscript k which is a non-zero multiple of p^{r_n} . If no such value exists either all the forms vanish, which occurs if and only if $x_1 = 0, x_2 = 0, \dots, x_n = 0$, or by Condition 1 at least one of the forms must take a non-zero integral value which we have assumed not to be a multiple of p^{r_n} ; thus if no such value L_k exists the truth of the theorem must be admitted. If we replace x_k by $x_k - L_k p^{-r_k}$ the forms $L_1(x), L_2(x), \dots, L_{k-1}(x)$ retain their original values while $L_k(x)$ takes the value zero. The remaining forms $L_{k+1}(x), L_{k+2}(x), \dots, L_n(x)$ take values which differ from the original values by multiples of p^{r_n} for by Theorem 1 $p^{-r_k}a_{jk}$ is integral for $j \geq k$ and we are assuming $p^{r_n} | L_k$. By repeating this process we ultimately replace x_1, x_2, \dots, x_n by a new set of integral values which give the forms

² B. L. van der Waerden, *Moderne Algebra*, § 106.

values which we may call L'_1, L'_2, \dots, L'_n . Each of this latter set of values differs by a multiple of p^{r_n} from the corresponding value of the set L_1, L_2, \dots, L_n . None of the values L'_1, L'_2, \dots, L'_n can be a non-zero multiple of p^{r_n} . In case all these values are zero we deduce L_1, L_2, \dots, L_n are all integral multiples of p^{r_n} . If on the other hand one of L'_1, L'_2, \dots, L'_n is different from zero then by Condition 1 at least one value say L'_h is a non-zero integer and furthermore an integer which is not a multiple of p^{r_n} . Consequently L_h is an integer which is not a multiple of p^{r_n} . The theorem is then completely established.

Form 2. r_1, r_2, \dots, r_n are integers not all of which are zero for which $r_1 \leq r_2 \leq \dots \leq r_n$ and $r_1 + r_2 + \dots + r_n = 0$. \mathfrak{F} is an Abelian group of rank n with type $(p^{-r_1+r_n}, p^{-r_2+r_n}, \dots, p^{-r_n+r_n})$; \mathfrak{S} a subgroup of \mathfrak{F} of type $(p^{-r_1+r_1}, p^{-r_2+r_2}, \dots, p^{-r_n+r_n})$; $\mathfrak{Z}_1, \mathfrak{Z}_2, \dots, \mathfrak{Z}_n$ a system of cyclic subgroups of \mathfrak{F} which together generate \mathfrak{F} . For every subgroup \mathfrak{A} of \mathfrak{F} containing \mathfrak{S} with $\mathfrak{F}/\mathfrak{A}$ cyclic it is known that a subgroup \mathfrak{Z}_r exists with $\mathfrak{A} \supseteq \mathfrak{Z}_r^{p^{r_n}}, \mathfrak{A} \not\supseteq \mathfrak{Z}_r$.

Then a subgroup \mathfrak{Z}_r exists with $\mathfrak{S} \supseteq \mathfrak{Z}_r^{p^{r_n}}, \mathfrak{S} \not\supseteq \mathfrak{Z}_r^{p^{r_n-1}}$.

Proof of equivalence to Conjecture. We shall first show how Form 1 of the Conjecture may be stated in terms of finite Abelian p -groups. Let $L_1(x), L_2(x), \dots, L_n(x)$ be a set of linear forms satisfying Conditions 1 and 2 and having the form of Theorem 1. We shall further assume that $r_1 < 0$. For otherwise from the conditions stated in Theorem 1 regarding the integers r_1, r_2, \dots, r_n we deduce that the coefficients of the forms are all integral in which case there is nothing to prove. This last assumption implies $n > 1$.

As x_1, x_2, \dots, x_n each independent of the other take all integral values modulo $p^{-r_1+r_n}$ let \mathfrak{G} be the group of vectors $p^{r_1}(x_1, x_2, \dots, x_n)$; \mathfrak{S} the subgroup of all vectors $(L_1(x), L_2(x), \dots, L_n(x))$; $\mathfrak{Z}_r, \mathfrak{Z}_r', \mathfrak{Z}_r''$ the subgroups of vectors

$$\begin{aligned} & p^{r_1}(x_1, \dots, x_{r-1}, 0, x_{r+1}, \dots, x_n), \\ & p^{r_1}(x_1, \dots, x_{r-1}, p^{-r_1+1}x_r, x_{r+1}, \dots, x_n), \quad 1 \leq r \leq n. \\ & p^{r_1}(x_1, \dots, x_{r-1}, p^{-r_1}x_r, x_{r+1}, \dots, x_n). \end{aligned}$$

As the forms have the form of Theorem 1, \mathfrak{S} has type $(p^{-r_1+r_n}, p^{-r_2+r_n}, \dots, p^{-r_n+r_n})$. From Theorem 2 follows that for every non-zero element A of \mathfrak{S} groups $\mathfrak{Z}_r, \mathfrak{Z}_r''$ exist with $A \in \mathfrak{Z}_r'', A \notin \mathfrak{Z}_r$ and conversely if the groups have this latter property the forms satisfy the condition of Theorem 2, which implies Condition 1. If one of the forms $L_r(x)$ has integral coefficients $\mathfrak{S} \subseteq \mathfrak{Z}_r''$ and conversely the existence of such a group \mathfrak{Z}_r'' implies that the corresponding form has integral coefficients. If the integral coefficients of $L_r(x)$ have no common factor

$\tilde{\mathfrak{S}} \not\subseteq \tilde{\mathfrak{Z}}_r'$. Again the converse is true for the relation $\tilde{\mathfrak{S}} \not\subseteq \tilde{\mathfrak{Z}}_r'$ implies that the integral coefficients of $L_r(x)$ are not all divisible by p ; because of the unit determinant and of the denominators of the forms none of them may have any other common factor.

From a system of linear forms we have constructed an equivalent system of Abelian groups in terms of which the Conjecture was restated. It is possible to start with an abstract system of vector groups $\mathfrak{G}, \tilde{\mathfrak{S}}, \tilde{\mathfrak{Z}}_r, \tilde{\mathfrak{Z}}_r', \tilde{\mathfrak{Z}}_r'', 1 \leq r \leq n$, with all the above relationships including the fact that $\tilde{\mathfrak{S}}$ have type $(p^{-r_1+r_n} p^{-r_2+r_n} \cdots p^{-r_n+r_n})$ where $r_1 + r_2 + \cdots + r_n = 0$ and $r_1 \leq r_2 \leq \cdots \leq r_n$ and by considering an automorphism of \mathfrak{G} on $\tilde{\mathfrak{S}}$ to construct a system of linear forms with unit determinant satisfying Conditions 1 and 2. In other words to every set of groups satisfying the conditions of the above group form of the Conjecture a set of linear forms exists satisfying the original conditions of the Conjecture. Thus the complete equivalence of the two forms is established.

Let \mathfrak{F} be the multiplicatively written character group of the group \mathfrak{G} . To every subgroup \mathfrak{B} of \mathfrak{G} we order the subgroup \mathfrak{B} of \mathfrak{F} of all characters χ for which $\chi(A) = 1$ for $A \in \mathfrak{B}$. It is a classical result of Weber that $\mathfrak{F} \cong \mathfrak{G}$ and that \mathfrak{B} determines \mathfrak{B} while $\mathfrak{F}/\mathfrak{B} \cong \tilde{\mathfrak{B}}$. Accordingly if $\tilde{\mathfrak{S}}$ be the character subgroup associated with $\tilde{\mathfrak{S}}$ it will have type $(p^{-r_1+r_1} p^{-r_2+r_2} \cdots p^{-r_n+r_n})$. The dual groups \mathfrak{Z}_r of the system $\tilde{\mathfrak{Z}}_r$ form a system of cyclic groups $\mathfrak{Z}_r = (A_r)$, $1 \leq r \leq n$, which generate the group \mathfrak{F} while the dual groups of the system $\tilde{\mathfrak{Z}}_r', \tilde{\mathfrak{Z}}_r''$ are the cyclic groups $(A_r p^{r_n-1}), (A_r p^{r_n})$, $1 \leq r \leq n$ respectively. Now if A be an element of $\tilde{\mathfrak{S}}$ the dual of the cyclic group (A) will be a subgroup \mathfrak{A} of \mathfrak{F} containing $\tilde{\mathfrak{S}}$ for which $\mathfrak{F}/\mathfrak{A}$ is cyclic. Conversely every subgroup \mathfrak{A} of \mathfrak{F} containing $\tilde{\mathfrak{S}}$, for which $\mathfrak{F}/\mathfrak{A}$ is cyclic, is the dual of a cyclic subgroup (A) of $\tilde{\mathfrak{S}}$. Accordingly, translating the conditions of the above group form of the Conjecture into the character groups, we see for every such subgroup \mathfrak{A} a subgroup \mathfrak{Z}_r exists with $\mathfrak{A} \supseteq \mathfrak{Z}_r p^{r_n}$, $\mathfrak{A} \not\supseteq \mathfrak{Z}_r$. The Conjecture itself becomes under similar translation: a subgroup \mathfrak{Z}_r exists with $\tilde{\mathfrak{S}} \supseteq \mathfrak{Z}_r p^{r_n}$, $\tilde{\mathfrak{S}} \not\supseteq \mathfrak{Z}_r p^{r_n-1}$. Thus the Conjecture in Form 1 is shown to be equivalent to Form 2 and the proof is complete.

Definition. For an Abelian group \mathfrak{G} of order p^{rn} a series of subgroups $\mathfrak{G} = \mathfrak{G}_1, \mathfrak{G}_2, \cdots, \mathfrak{G}_s, \mathfrak{G}_{s+1} = (E)$ is said to form an r -series if $\mathfrak{G}_s/\mathfrak{G}_{s+1}$ is cyclic and of order p^r for $1 \leq s \leq n$.

Definition. A subgroup \mathfrak{S} of \mathfrak{G} is said to be reciprocally cyclic if the factor group $\mathfrak{G}/\mathfrak{S}$ is cyclic.

Form 3. \mathfrak{B} is an Abelian group of order p^{rn} with rank less than n .

$\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ are n cyclic subgroups of \mathfrak{B} . For every reciprocally cyclic subgroup \mathfrak{D} of \mathfrak{B} groups $\mathfrak{C}_{s_1}, \mathfrak{C}_{s_2}, \dots, \mathfrak{C}_{s_k}$ exist so that the factors of the series

$$\mathfrak{B} = (\mathfrak{C}_{s_1}, \mathfrak{C}_{s_2}, \dots, \mathfrak{C}_{s_k}, \mathfrak{D}), \dots, (\mathfrak{C}_{s_1}, \mathfrak{D}), \mathfrak{D}$$

have order not greater than p^r . Then the groups $\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ after a possible rearrangement of order build an r -series

$$\mathfrak{B} = (\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n), \dots, (\mathfrak{C}_1, \mathfrak{C}_2), \mathfrak{C}_1, (E)$$

for the group \mathfrak{B} .

Proof of equivalence to Conjecture. We first show the above would follow from a proof of the Conjecture expressed in Form 2. $\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ together generate the group \mathfrak{B} for otherwise a group \mathfrak{D} containing $\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ would exist for which $\mathfrak{B}/\mathfrak{D}$ was cyclic and of order greater than 1. Then by hypothesis a group \mathfrak{C}_t would exist for which $(\mathfrak{C}_t, \mathfrak{D}) : \mathfrak{D}$ would be greater than 1, contradicting the fact that $\mathfrak{C}_t \leq \mathfrak{D}$.

Let r_1, r_2, \dots, r_n be integers for which $r_1 \leq r_2 \leq \dots \leq r_n$ and such that \mathfrak{B} has type $(p^{-r_1+r} p^{-r_2+r} \dots p^{-r_n+r})$. Now $r_n = r$ because the rank of \mathfrak{B} is by hypothesis less than n . Furthermore the order of \mathfrak{B} being p^{rnn} , $r_1 + r_2 + \dots + r_n = 0$. Now let \mathfrak{F} be a group of type $(p^{-r_1+r_n} p^{-r_2+r_n} \dots p^{-r_1+r_n})$ of rank n generated by elements Z_1, Z_2, \dots, Z_n . If C_1, C_2, \dots, C_n be elements of \mathfrak{B} which generate the cyclic subgroups $\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ respectively, we define a homomorphism of \mathfrak{F} on \mathfrak{B} by the correspondence $Z_1 \rightarrow C_1, Z_2 \rightarrow C_2, \dots, Z_n \rightarrow C_n$. This is possible because C_1, C_2, \dots, C_n generate \mathfrak{B} and the order of each element of \mathfrak{F} is a multiple of the order of corresponding element of \mathfrak{B} . Let \mathfrak{S} be the subgroup of \mathfrak{F} which is built into the unit element of \mathfrak{B} in the above homomorphism. As $\mathfrak{F}/\mathfrak{S} \cong \mathfrak{B}$ we deduce from the type of \mathfrak{F} and \mathfrak{B} that \mathfrak{S} has type $(p^{-r_1+r_1} p^{-r_2+r_2} \dots p^{-r_1+r_n})$. For a subgroup \mathfrak{A} of \mathfrak{F} containing \mathfrak{S} we have by the second isomorphism theorem $\mathfrak{F}/\mathfrak{A} \cong \mathfrak{F}/\mathfrak{S}/\mathfrak{A}/\mathfrak{S}$. Hence if \mathfrak{A} is reciprocally cyclic, \mathfrak{A} is built by the homomorphism into a reciprocally cyclic subgroup \mathfrak{D} of \mathfrak{B} . Now by the hypothesis, as $r_n = r$, a subgroup \mathfrak{C}_s exists with $1 < ((\mathfrak{C}_s, \mathfrak{D}) : \mathfrak{D}) \leq p^{r_n}$ from which we can conclude $\mathfrak{A} \cong (Z_s^{p^{r_n}})$, $\mathfrak{A} \not\cong (Z_s)$. Thus \mathfrak{F} and its subgroups (Z_r) , $1 \leq r \leq n$, satisfy all the conditions of Form 2. Therefore if the Conjecture be true a number s_1 exists with $\mathfrak{S} \cong (Z_{s_1}^{p^{r_n}})$, $\mathfrak{S} \not\cong (Z_{s_1}^{p^{r_{n-1}}})$. This implies $\mathfrak{C}_{s_1}^{p^{r_n}} = (E)$, $\mathfrak{C}_{s_1}^{p^{r_{n-1}}} \neq (E)$ i. e. \mathfrak{C}_{s_1} has exact order p^{r_n} . But as $r_n = r$, \mathfrak{C}_{s_1} has order p^r .

In the factor group $\mathfrak{B}/\mathfrak{C}_{s_1}$ let $\mathfrak{C}'_1, \dots, \mathfrak{C}'_{s_1-1}, \mathfrak{C}'_{s_1+1}, \dots, \mathfrak{C}'_n$ be the cyclic subgroups of restclasses defined by $\mathfrak{C}_1, \dots, \mathfrak{C}_{s_1-1}, \mathfrak{C}_{s_1+1}, \dots, \mathfrak{C}_n$ respectively. By using the second isomorphism theorem, any reciprocally cyclic subgroup \mathfrak{D}'

of $\mathfrak{B}/\mathfrak{C}_{s_1}$ may be shown to have the form $\mathfrak{D}/\mathfrak{C}_{s_1}$ where \mathfrak{D} is a reciprocally cyclic subgroup of \mathfrak{B} . By hypothesis a series

$$(\mathfrak{C}_{t_1}, \dots, \mathfrak{C}_{t_k}, \mathfrak{D}), \dots, (\mathfrak{C}_{t_1}, \mathfrak{D}), \mathfrak{D}$$

exists from \mathfrak{B} to \mathfrak{D} whose factors have order not greater than p^r . As $\mathfrak{D} \geq \mathfrak{C}_{s_1}$ it follows from the second isomorphism theorem that the factors of this series are isomorphic to the factors of the series

$$\mathfrak{B}/\mathfrak{C}_{s_1} = (\mathfrak{C}'_{t_1}, \dots, \mathfrak{C}'_{t_k}, \mathfrak{D}'), \dots, (\mathfrak{C}'_{t_1}, \mathfrak{D}'), \mathfrak{D}'.$$

Therefore the factors of this series have order not greater than p^r . We have proved the order of \mathfrak{C}_{s_1} is p^r . Hence the order of the factor group $\mathfrak{B}/\mathfrak{C}_{s_1}$ is $p^{n(r-1)}$. We have thus proved that the factor group $\mathfrak{B}/\mathfrak{C}_{s_1}$ and its associated subgroups $\mathfrak{C}'_1, \dots, \mathfrak{C}'_{s_1-1}, \mathfrak{C}'_{s_1+1}, \dots, \mathfrak{C}'_n$ satisfy all the conditions of Form 3 with n replaced by $n-1$. We could therefore deduce from the truth of the Conjecture exactly as above that at least one element \mathfrak{C}'_{s_2} has order p^r which means $(\mathfrak{C}_{s_1}, \mathfrak{C}_{s_2}) : \mathfrak{C}_{s_1} = p^r$.

By repeating this process we deduce after n steps that $\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ after a possible rearrangement of order build an r -series

$$(\mathfrak{C}_1, \dots, \mathfrak{C}_n), \dots, (\mathfrak{C}_1, \mathfrak{C}_2), \mathfrak{C}_1, (E)$$

for \mathfrak{B} . This shows that the problem stated in Form 3 is a consequence of the Conjecture as stated in Form 2.

To show the converse we need only consider the factor group $\mathfrak{F}/\mathfrak{S}$ and its associated cyclic subgroups of restclasses $\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ defined by $\mathfrak{Z}_1, \mathfrak{Z}_2, \dots, \mathfrak{Z}_n$ respectively. $\mathfrak{F}/\mathfrak{S}$ has order p^{nrn} . For any reciprocally cyclic \mathfrak{D} of $\mathfrak{F}/\mathfrak{S}$ let \mathfrak{D}' be the subgroup of \mathfrak{F} of all elements which are built into \mathfrak{D} by the homomorphism $\mathfrak{F} \simeq \mathfrak{F}/\mathfrak{S}$. \mathfrak{D}' by the second isomorphism theorem is reciprocally cyclic, hence by the hypothesis of Form 2 a subgroup \mathfrak{Z}_{t_1} exists with $\mathfrak{D}' \geq \mathfrak{Z}_{t_1}^{p^r n}$, $\mathfrak{D}' \not\geq \mathfrak{Z}_{t_1}$. Therefore $(\mathfrak{D}, \mathfrak{C}_{t_1})/\mathfrak{D}$ has order greater than 1 but not greater than $p^r n$. Now the subgroup $(\mathfrak{D}, \mathfrak{C}_{t_1})$ is also reciprocally cyclic and so proceeding exactly as before we could find a subgroup \mathfrak{C}_{t_2} with $(\mathfrak{C}_{t_1}, \mathfrak{C}_{t_2}, \mathfrak{D})/(\mathfrak{C}_{t_1}, \mathfrak{D})$ of order greater than 1 but not greater than $p^r n$. Continuing in this manner in a finite number of steps we could construct a series

$$\mathfrak{F}/\mathfrak{S} = (\mathfrak{C}_{t_1}, \dots, \mathfrak{C}_{t_k}, \mathfrak{D}), \dots, (\mathfrak{C}_{t_1}, \mathfrak{D}), \mathfrak{D}$$

with cyclic factors of order not greater $p^r n$. Thus $\mathfrak{F}/\mathfrak{S}$ and its subgroups $\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ are seen to satisfy all the conditions of Form 3 with r replaced by r_n . Therefore if Form 3 of the Conjecture were true a subgroup \mathfrak{C}_s would exist with exact order $p^r n$ which would imply: $\mathfrak{S} \geq \mathfrak{Z}_s^{p^r n}$, $\mathfrak{S} \not\geq \mathfrak{Z}_s^{p^r n-1}$ which is Form 2 of the Conjecture. Therefore Forms 2 and 3 are completely equivalent.

REMARKS ON MULTIGROUPS.*

By J. E. EATON and OYSTEIN ORE.

The present paper may be considered as a supplement to a recent paper on multigroups by Dresher and Ore.¹ It contains various contributions which lead to simplifications and improvements in certain parts of the previous theory of multigroups. The notations and terminology are the same and need therefore not be explained here.

1. Existence of cross-cut. In the following let \mathfrak{M} denote a multigroup and let \mathfrak{A} and \mathfrak{B} be submultigroups. The theory of submultigroups differs from ordinary group theory in that the cross-cut $(\mathfrak{A}, \mathfrak{B})$ may be void. On the other hand certain theorems in the theory of normal submultigroups require that the cross-cut of particular submultigroups shall not be void; hence this must be stated as a separate condition, or it can be fulfilled by assuming that the multigroup contains units. We shall show however that in some of the most important cases these conditions are not necessary because the existence of a cross-cut follows from:

THEOREM 1. *Let \mathfrak{A} be a left reversible and \mathfrak{B} a left closed submultigroup of \mathfrak{M} . Then the cross-cut $(\mathfrak{A}, \mathfrak{B})$ is not void.*

Proof. Let a and b be elements in \mathfrak{A} and \mathfrak{B} respectively and let us determine m such that

$$am \supset b.$$

By the reversibility of \mathfrak{A} follows $m \subset a_1b$ or

$$aa_1b \supset a_2b \supset b$$

and here a_2 must belong to \mathfrak{B} since \mathfrak{B} is left closed.

From Theorem 1 follows further:

THEOREM 2. *Let \mathfrak{A} be normal and left reversible while \mathfrak{B} is left closed in \mathfrak{M} . Then every element in the union $[\mathfrak{A}, \mathfrak{B}]$ is contained in a product ab .*

Proof. It is obvious from the definition of normality that any element not in \mathfrak{A} or \mathfrak{B} must be contained in such a product and for the elements in

* Received April 25, 1939.

¹ Melvin Dresher and Oystein Ore, "Theory of multigroups," *American Journal of Mathematics*, vol. 60 (1938), pp. 705-733. We shall quote this paper in the following as D. and O.

\mathfrak{A} and \mathfrak{B} it follows from the existence of an element d belonging both to \mathfrak{A} and \mathfrak{B} .

On the basis of these two theorems one obtains the main properties of normal submultigroups:²

Let \mathfrak{A} and \mathfrak{B} be normal reversible submultigroups. Then the union $[\mathfrak{A}, \mathfrak{B}]$ is also normal and reversible and the cross-cut $(\mathfrak{A}, \mathfrak{B})$ is normal and reversible in \mathfrak{A} and in \mathfrak{B} .

If \mathfrak{A} is normal and reversible and \mathfrak{B} closed in the union $[\mathfrak{A}, \mathfrak{B}]$ then $(\mathfrak{A}, \mathfrak{B})$ is normal and reversible in \mathfrak{B} and there exists an isomorphism between the quotient systems.

$$[\mathfrak{A}, \mathfrak{B}]/\mathfrak{A} \cong \mathfrak{B}/(\mathfrak{A}, \mathfrak{B}).$$

2. Homomorphisms. A multigroup \mathfrak{M} is said to be homomorphic to another \mathfrak{M}^* when there exists a correspondence $m \rightarrow m^*$ between their elements such that

$$ab \supset c$$

implies

$$a^*b^* \supset c^*.$$

Furthermore every element of \mathfrak{M}^* shall be the image of some element of \mathfrak{M} .

One proves (D. and O. Theorem 12, Chapter 2) that if \mathfrak{M}^* contains a left scalar unit e^* then all elements \mathfrak{A} of \mathfrak{M} corresponding to e^* form a right multigroup which is right closed and if e^* is an absolute unit element of \mathfrak{M}^* then \mathfrak{A} is a closed submultigroup.

In order to derive further properties of the homomorphism it is necessary to make assumptions on the inverse correspondence from \mathfrak{M}^* to \mathfrak{M} .

We shall say that \mathfrak{M} is *left properly homomorphic* to \mathfrak{M}^* when:

1. \mathfrak{M}^* contains a left scalar unit e^* .

We denote by \mathfrak{A} the right closed right multigroup consisting of the elements in \mathfrak{M} corresponding to e^* .

2. If $m^*_1 = m^*_2$ then there exist elements a_1 and a_2 in \mathfrak{A} such that

$$a_1 m_1 \supset m_2, \quad a_2 m_2 \supset m_1.$$

This condition shows that \mathfrak{A} is left reversible. Hence there exists a coset expansion of \mathfrak{M} with respect to \mathfrak{A} (D. and O. Theorem 9, Chapter 2) and it is easily shown:

THEOREM 3. *If a multigroup \mathfrak{M} is left properly homomorphic to another \mathfrak{M}^* then \mathfrak{M}^* is isomorphic to a quotient multigroup*

² These are somewhat simplified statements of the results in D. and O., chap. 3, § 2.

$$\mathfrak{M}^* \cong \mathfrak{M}/\mathfrak{A}$$

where \mathfrak{A} is a left reversible right submultigroup of \mathfrak{M} .³

In the paper by Dresher and Ore also the following type of homomorphism has been introduced: A multigroup \mathfrak{M} is *strongly (left) homomorphic* to \mathfrak{M}^* when any relation

$$a^*b^* \supset c^*$$

implies that to any b_0 and c_0 corresponding to b^* and c^* respectively there exists some a corresponding to a^* such that

$$ab_0 \supset c_0.$$

One can then show:

THEOREM 4. *Let \mathfrak{M} be (left and right) properly homomorphic to \mathfrak{M}^* . Then \mathfrak{M} is also strongly homomorphic to \mathfrak{M}^* .*

Proof. When \mathfrak{M} is both left and right properly homomorphic to \mathfrak{M}^* there must exist an absolute unit e^* in \mathfrak{M}^* and those elements in \mathfrak{M} which correspond to e^* form a reversible submultigroup \mathfrak{A} . But in this case \mathfrak{A} must also be normal since all m with the same image m^* must be contained both in a coset $m\mathfrak{A}$ and a coset $\mathfrak{A}m$. From Theorem 3 it follows that \mathfrak{M}^* is isomorphic to the quotient multigroup $\mathfrak{M}/\mathfrak{A}$. To show that \mathfrak{M} is strongly homomorphic to $\mathfrak{M}/\mathfrak{A}$ let us assume that a relation

$$(1) \quad m_1\mathfrak{A} \cdot m_2\mathfrak{A} \supset m_3\mathfrak{A}$$

holds for three cosets. Any m corresponding to $m_1\mathfrak{A}$ may be taken as the multiplier of this coset and similarly for $m_3\mathfrak{A}$. Hence we shall only have to show that (1) implies the existence of some element x in $m_2\mathfrak{A}$ such that

$$m_1x \supset m_3$$

and this follows from the normality of \mathfrak{A} .

This also implies D. and O. Theorem 1, Chapter 3.

3. Strong normality. An important concept in the theory of multigroups is that of *strong normality*. This concept has been defined (D. and O. §4, Chapter 3) under the assumption that right and left units and hence inverses exist in the multigroup \mathfrak{M} . We shall show here that one can also give alternative definitions in which these assumptions are not necessary.

DEFINITION. *A closed submultigroup \mathfrak{A} of \mathfrak{M} is strongly normal if for any m_1 there exists an m'_1 such that*

³ This theorem is a corrected form of Theorem 13, chap. 2 in D. and O. In all statements on p. 721 strong homomorphism should be replaced by proper homomorphism.

$$(2) \quad \mathfrak{A} \supset m_1 \mathfrak{A} m'_1$$

and to every m_2 an m'_2 such that

$$(3) \quad \mathfrak{A} \supset m'_2 \mathfrak{A} m_2.$$

Let us derive some consequences of this definition. We determine m_1'' such that

$$(4) \quad \mathfrak{A} \supset m'_1 \mathfrak{A} m_1''$$

and from (2) one obtains

$$(5) \quad m_1 \mathfrak{A} m'_1 \mathfrak{A} m_1'' = \mathfrak{A} m_1'' = m_1 \mathfrak{A}.$$

When this is substituted in (4) one finds further

$$\mathfrak{A} \supset m'_1 m_1 \mathfrak{A}$$

and since \mathfrak{A} is closed

$$m'_1 m_1 \subset \mathfrak{A}.$$

Similarly one has

$$m_2 m'_2 \subset \mathfrak{A}.$$

We show next:

A strongly normal submultigroup is reversible.

If namely

$$a_1 m_1 \supset m_2$$

then one can determine some z such that

$$z m_2 \supset m_1$$

or

$$a_1 z m_2 \supset a_1 m_1 \supset m_2.$$

When this relation is multiplied by m'_2 one finds a relation of the form

$$a_1 z a_2 \supset a_3$$

showing that z belongs to \mathfrak{A} since \mathfrak{A} is closed.

A strongly normal submultigroup is normal.

From the reversibility it follows that right and left coset expansions of \mathfrak{M} with respect to \mathfrak{A} exist and since each coset contains its multiplier we obtain from (5)

$$\mathfrak{A} m_1 = m_1 \mathfrak{A}.$$

Let us finally consider the quotient multigroup $\mathfrak{M}/\mathfrak{A}$. From the condition of strong normality it follows that for any m_1 and m_2 one can find an m_3 such that

$$\mathfrak{A}m_3 \supset \mathfrak{A}m_1 \cdot \mathfrak{A}m_2.$$

This indicates however that the product of two cosets contains only a single coset; hence the multigroup $\mathfrak{M}/\mathfrak{A}$ is an ordinary group.

THEOREM 5. *The necessary and sufficient condition that a multigroup \mathfrak{M} be homomorphic to a group \mathfrak{G} is that \mathfrak{M} contain a strongly normal submultigroup \mathfrak{A} such that*

$$\mathfrak{G} \cong \mathfrak{M}/\mathfrak{A}.$$

We have already proved the sufficiency of this condition and the necessity follows by the same argument as in the proof of Theorem 12, Chapter 3 in D. and O.

Let us prove finally:

THEOREM 6. *The strongly normal submultigroups form a Dedekind structure.*

Proof. Since the Dedekind relation holds for normal submultigroups (D. and O. Theorem 5, Chapter 3) we shall only have to show that the submultigroups form a structure.

The union of two strongly normal submultigroups \mathfrak{A} and \mathfrak{B} is closed (D. and O. Theorem 5, Chapter 2) and $[\mathfrak{A}, \mathfrak{B}] = \mathfrak{A}\mathfrak{B}$ (Theorem 2). To any m let m' be determined such that $mm' \subset \mathfrak{A}$. Then

$$m\mathfrak{A}\mathfrak{B}m' = mm'\mathfrak{A}\mathfrak{B} = \mathfrak{A}\mathfrak{B}.$$

The cross-cut $\mathfrak{D} = (\mathfrak{A}, \mathfrak{B})$ is also closed (D. and O. Theorem 4, Chapter 2). Let m and m' be arbitrary. Then

$$\mathfrak{A}x \supset m\mathfrak{D}m', \quad \mathfrak{B}x \supset m\mathfrak{D}m'$$

for any x in mm' . If m' is determined such that $mm' \supset d$, where \mathfrak{D} contains d , then

$$\mathfrak{A} \supset m\mathfrak{D}m', \quad \mathfrak{B} \supset m\mathfrak{D}m'$$

and hence

$$\mathfrak{D} \supset m\mathfrak{D}m'.$$

Theorem 6 again implies the existence of a unique minimal strongly normal submultigroup \mathfrak{A}_0 such that $\mathfrak{M}/\mathfrak{A}_0$ is a group.

To conclude let us remark that a submultigroup \mathfrak{A} can also be said to be strongly normal if it is left reversible and the relation (2) holds. This definition can be shown to be equivalent to the preceding.

ON THE IMBEDDING OF ONE SEMI-GROUP IN ANOTHER, WITH APPLICATION TO SEMI-RINGS.*

By H. S. VANDIVER.

In other papers¹ a semi-group was defined as a set of elements closed under an associative operation and for which the equivalence and the substitution postulates hold. In the present paper we shall employ instead of the substitution postulate, the postulate that if $A = B$, then $CA = CB$ and $AC = BC$ for any A, B or C in the set, which we shall call the composition postulate.²

A gruppoid³ is a semi-group with an identity element, that is, an E such that, $AE = EA = A$ for any A in the set. A quasi-group is a semi-group such that from either of the relations

and

$$AB = AC$$

$$BD = CD$$

we infer

$$B = C$$

where each letter denotes an element of the semi-group. Cancellable elements in a semi-group S are elements C such that if $CM = CN$ then $M = N$ or if $HC = KC$ then $H = K$, where each letter denotes an element of S . It is easy to see that the product of two cancellable elements in S is also cancellable; hence these elements form a sub-set of S which is a quasi-group. Isomorphism between two semi-groups is defined in the same way as for the isomorphism between two groups, and the *central* of S will be the set of elements in S which are permutable with each element of S , as in group theory. A semi-group S will be said to be imbedded (or immersed) in another semi-group S' if S' contains a sub-semi-group which is isomorphic to S .

Graves⁴ in a recent paper showed how to immerse a commutative quasi-

* Received July 3, 1939.

¹ Vandiver, *Proceedings of the National Academy of Sciences*, vol. 20 (1934), p. 579; *Bulletin of the American Mathematical Society*, vol. 40 (1934), p. 916; *American Mathematical Monthly*, vol. 46 (1939), p. 24.

² The relations between these two sets of postulates and to other sets of postulates for a semi-group I hope to discuss elsewhere.

³ Here we follow the terminology used by Specht and Garrett Birkhoff. Cf. the latter, *Annals of Mathematics*, vol. 35 (1934), p. 351 and note references there given.

⁴ *American Mathematical Monthly*, vol. 45 (1938), pp. 664-69. Graves calls the system I have called a quasi-group a semi-group.

group in a group. In the present paper we prove a generalization (Th. 1) of this result and apply it to the theory of semi-rings.

THEOREM 1. *If a semi-group S contains a cancellable element, and all the cancellable elements of S belong to its central, then S may be imbedded in a groupoid S' whose cancellable elements form an Abelian group G , and the identity element of G is the identity element of S' .*

For proof, let the distinct elements of S be denoted by

$$a_1, a_2, \dots$$

and in particular the distinct elements of this set which are cancellable by

$$c_1, c_2, \dots$$

since the latter, by hypothesis is not a null-set.

Then consider the set S' formed by the pairs

$$(a_i, c_s)$$

and write

$$(3) \quad (a_i, c_s) (a_k, c_t) = (a_i a_k, c_s c_t).$$

Also we shall agree that

$$(4) \quad (a_h, c_r) = (a_i, c_v)$$

if and only if

$$a_h c_v = a_i c_r.$$

By (3) the closure law holds for the pairs since the closure holds for the a 's, and also for the c 's. We now examine the equivalence postulates.

$$(a_i, c_s) = (a_i, c_s)$$

obviously holds from (4). Symmetry obviously holds since it holds for the a 's. As for transitivity, if

$$(a_i, c_s) = (a_j, c_t)$$

and

$$(a_j, c_t) = (a_k, c_r)$$

then

$$(5) \quad a_i c_t = a_j c_s, \quad a_j c_r = a_k c_t;$$

whence, since composition holds in S ,

$$a_i c_t c_r = a_j c_s c_r$$

and if the c 's belong to the central of S then

$$a_i c_r c_t = a_j c_r c_s$$

or from (5), using composition and transitivity in S ,

$$a_i c_r c_t = a_j c_r c_s = a_k c_t c_s = a_k c_s c_t$$

and since c_t is a cancellable element in S then

$$a_i c_r = a_k c_s$$

and by (4)

$$(a_i, c_s) = (a_k, c_r).$$

Transitivity then holds. We also have that if

$$(6) \quad (a_i, c_s) = (a_k, c_r)$$

then

$$(7) \quad (a_i, c_s) (a_j, c_t) = (a_k, c_r) (a_j, c_t),$$

for we have from (6)

$$a_i c_r = a_k c_s$$

and

$$a_i c_r a_j c_t = a_k c_s a_j c_t$$

or

$$(a_i a_j, c_s c_t) = (a_k a_j, c_r c_t);$$

whence we obtain (7), using (3) and (4). Hence composition holds since (7) holds for the reverse order of the factors. It is easy to verify that the associative law holds for the pairs, hence, with the above conclusions, we have proved that S' is a semi-group. We shall now show that S' is a groupoid. Since S by hypothesis contains a cancellable element, say c , we note that

$$(8) \quad (a_i, c_s) (c, c) = (a_i c, c_s c)$$

and also

$$a_i c_s c = a_i c_s c = a_i c c_s,$$

whence

$$(a_i c, c_s c) = (a_i, c_s)$$

and (8) gives

$$(a_i, c_s) (c, c) = (a_i, c_s).$$

We find similarly that

$$(c, c) (a_i, c_s) = (a_i, c_s)$$

so that (c, c) is an identity element of S' , and S' is then a groupoid.

We shall now show that all elements of the form

$$(c_s, c_t)$$

are cancellable in S' . For suppose that

$$(c_s, c_t) (a_i, c_r) = (c_s, c_t) (a_j, c_v);$$

then

$$(c_s a_i, c_t c_r) = (c_s a_j, c_t c_v)$$

by (3), and (4) gives

$$c_s a_i c_t c_v = c_s a_j c_t c_r$$

and since the c 's belong to the central of S and are also cancellable in S we have

$$a_i c_v = a_j c_r,$$

or

$$(a_i, c_r) = (a_j, c_v).$$

We obtain a similar result for the other order of the factors.

Hence (c_s, c_t) is a cancellable element in S' .

It will now be proved that any element of the form

$$(a_i, c_s)$$

is non-cancellable in S' if a_i is non-cancellable in S . For in that case there exist elements a_j and a_k with $a_j \neq a_k$ and

$$(9) \quad a_j a_i = a_k a_i,$$

or there exist elements a_{j_1} and a_{k_1} with $a_{k_1} \neq a_{j_1}$ and such that

$$(10) \quad a_i a_{j_1} = a_i a_{k_1}.$$

Now if (9) holds we have

$$(a_j a_i, c c_s) = (a_k a_i, c c_s)$$

or

$$(a_j, c) (a_i, c_s) = (a_k, c) (a_i, c_s)$$

with

$$(a_j, c) \neq (a_k, c).$$

We obtain the same result after treating (10) in a similar way.

Hence all the cancellable elements of S are of the form (c_s, c_t) . These form a group G in S' since for given elements (c_s, c_t) and (c_h, c_i) we may verify that

$$(c_s, c_t) (c_h c_t, c_s c_i) = (c_h, c_i)$$

and similarly for the other order of the factors on the left, and the result follows if we note the product of two cancellable elements in S is a cancellable

element in S . Also the group is Abelian since the c 's are commutative in S . We now show that

$$a_1, a_2, \dots$$

is isomorphic with

$$(a_1c, c), (a_2c, c), \dots$$

To show this we note first that the elements of the last set are distinct since the a 's are, for

$$(a_ic, c) = (a_jc, c)$$

gives

$$a_ic^2 = a_jc^2,$$

or

$$a_i = a_j.$$

Now if

$$(a_hc, c) \leftrightarrow a_h; h = 1, 2, 3, \dots$$

and

$$(a_ic, c)(a_jc, c) = (a_kc, c)$$

then

$$a_ia_j = a_k$$

and conversely; hence the sets are isomorphic. Also the identity element of G is (c, c) , so that the identity element of G is the identity element of S' and our theorem is proved.

We note the peculiarity that in order to prove the transitivity law in S' we make use of the fact that the c 's are cancellable in S . We may note that if we use a non-cancellable element n selected from the a 's in connection with the pairs of the type

$$(a, n)$$

then it is possible to select a particular set such that the law of transitivity does not hold within it. For let S contain an annihilator, say k , then if we use the definition of equality as in (3) we have

$$(k, k) = (a_i, c_j)$$

for any i and j , but transitivity does not hold since $(a_i, c_j) \neq (a_s, c_j)$ for $i \neq s$ and c_j a cancellable element. Since the pairs obviously obey laws similar to those which fractions follow under multiplication in ordinary arithmetic, we see that the above situation is similar to what we have in arithmetic when we attempt to employ the fraction $0/0$.

Application to semi-rings. Following closely another paper⁵ we define

⁵ *Proceedings of the National Academy of Sciences*, vol. 21 (1935), p. 162.

a semi-ring as a system of elements which form a semi-group under addition, a semi-group under multiplication and the right and left distributive laws hold. Let S now form a semi-ring and employ the notation

$$\frac{a}{c}$$

for (a, c) and define addition of these symbols by means of

$$(11) \quad \frac{a_1}{c_1} + \frac{a_2}{c_2} = \frac{a_1 c_2 + a_2 c_1}{c_1 c_2},$$

and call

$$\frac{a_1}{c_1} \cdot \frac{a_2}{c_2} = \frac{a_1 a_2}{c_1 c_2},$$

multiplication. It is easily seen that addition is associative since S is a semi-ring. The distributive law holds since

$$\begin{aligned} \frac{a_1}{c_1} \left(\frac{a_2}{c_2} + \frac{a_3}{c_3} \right) &= \frac{a_1}{c_1} \cdot \frac{a_2 c_3 + a_3 c_2}{c_2 c_3} \\ &= \frac{(a_1 a_2)(c_1 c_3) + (a_1 a_3)(c_1 c_2)}{(c_1 c_2)(c_1 c_3)}, \end{aligned}$$

since c is a cancellable element under multiplication in S . But the last expression on the right equals

$$\frac{a_1 a_2}{c_1 c_2} + \frac{a_1 a_3}{c_1 c_3},$$

and similarly we find

$$\left(\frac{a_2}{c_2} + \frac{a_3}{c_3} \right) \frac{a_1}{c_1} = \frac{a_2 a_1}{c_2 c_1} + \frac{a_3 a_1}{c_3 c_1}.$$

It is easily seen then, that S' , consisting of the elements a/c , is a semi-ring. Hence we have the

THEOREM 2. *A semi-ring R whose multiplicative semi-group S contains a cancellable element, and the cancellable elements of S belong to the central of S , may be imbedded in a semi-ring R' whose multiplicative semi-group is a gruppoid. The cancellable elements of this gruppoid form a group whose identity element is the identity of the gruppoid.*

We shall call R' the *quotient semi-ring* of R .

Suppose now that R is a ring, then since its additive semi-group is an

Abelian group using (11) we see that R' forms an additive Abelian semi-group since R does and it is a group since

$$\frac{a_1}{c_1} + x = \frac{a_2}{c_2}$$

has the solution

$$x = \frac{a_2 c_1 - a_1 c_2}{c_1 c_2}$$

and the word semi-ring may be replaced by ring in the statement of Theorem 2. Also we see that any non-cancellable element n of the multiplicative semi-group of a ring is zero or a zero divisor in the ring, since

$$na = nb$$

with $a \neq b$ gives

$$n(a - b) = 0;$$

and n is zero or a zero divisor since $a - b \neq 0$. A ring where division is always uniquely possible when the divisor is not zero or a zero divisor is called a quasi-field by the writer in another paper.⁶ This term however is employed in a different sense by other writers.⁷ If R is a realm or domain of integrity then R' is a field called the quotient field of R . This special case of Theorem 2 is well known.⁸

UNIVERSITY OF TEXAS.

⁶ *Proceedings of the National Academy of Sciences*, vol. 21 (1935), p. 162.

⁷ Cf., for example, Albert, *Modern Higher Algebra*, p. 23.

⁸ Van der Waerden, *Moderne Algebra*, 1st ed., Bd. 1, p. 7; Albert, *loc. cit.*, pp. 27-29.

NOTE ON EULER NUMBER CRITERIA FOR THE FIRST CASE OF FERMAT'S LAST THEOREM.*

By H. S. VANDIVER.

For the solution of

$$(1) \quad x^l + y^l + z^l = 0$$

$xyz \not\equiv 0 \pmod{l}$; x, y and z rational integers, l a given odd prime, we have the Kummer criteria

$$(2) \quad \begin{aligned} B_n f_{l-2n}(t) &\equiv 0 \pmod{l}, \\ f_{l-1}(t) &\equiv 0 \pmod{l} \end{aligned}$$

where

$$(2a) \quad \begin{aligned} -t &\equiv x/y, y/x, y/z, z/y, x/z, z/x, \text{ modulo } l, \\ &\quad (n = 1, 2, \dots, (l-3)/2), \\ f_a(w) &= \sum_{i=1}^{l-1} i^{a-1} w^i. \end{aligned}$$

Further, the B 's are the Bernoulli numbers, $B_1 = 1/6$, $B_2 = 1/30$, etc. Now all the known criteria which have been derived from (2) for the solution of (1) and which are independent of x, y , and z may be shown to have a certain relation to each other. All may be derived from a set of criteria of the form

$$(3) \quad C(m, i, r) = \sum_{s=[(r-1)l/m]+1}^{[rl/m]} s^i \equiv 0 \pmod{l}$$

for certain small values of m and where $i = l-2, l-3, l-4, l-6, l-8, l-10, l-12$; with r in the set $1, 2, \dots, m-1$, and criteria consisting of linear functions of the type (3) with rational coefficients.

It is known that¹

$$(4) \quad b_{2a} \frac{1-n^{2a}}{2an^{2a-1}} \equiv \sum_{i=1}^{[n/2]} (n-2i+1)C(n, 2a-1, i)$$

modulo l , for $l > 3$; $n \not\equiv 0 \pmod{l}$; $[h]$ is the greatest integer in h , and the b 's are defined by

$$(b+1)^n = b_n; \quad n > 1$$

where the left-hand member is expanded by the binomial theorem and b_4 substituted for b^4 . As is known $(-1)^{a-1}B_a = b_{2a}$. For $n = l-1$, (4) gives

$$b_{l-1} \cdot \frac{1-n^{l-1}}{(l-1)ln^{l-2}} \equiv \sum_{i=1}^{[n/2]} (n-2i+1)C(n, l-2, i)$$

* Received July 3, 1939.

¹ Vandiver, *Duke Mathematical Journal*, vol. 3 (1937), p. 572, relation 10.

modulo l , and using

$$-1 \equiv 1^{l-1} + 2^{l-1} + \cdots + (l-1)^{l-1} \equiv b_{l-1}l \pmod{l}$$

we have

$$(5) \quad \frac{1 - n^{l-1}}{ln^{l-2}} \equiv \sum_{i=1}^{[n/2]} (n - 2i + 1)C(n, l-2, i)$$

modulo l . Frobenius² found criteria of the type (3) for $i = l-2$, and various values of $m \leq 26$. Morishima,³ showed that if (1) is satisfied in Case 1 then

$$m^{l-1} \equiv 1 \pmod{l^2}$$

for each m such that $0 < m \leq 31$. Employing (5) we obtain criteria of the type mentioned above.

It has been shown, using (2), that for $2n = l-3, l-5, l-7, l-9, l-11, l-13$, we have the criteria $B_n \equiv 0 \pmod{l}$ in Case 1. Using (4) we obtain more criteria of the type mentioned in connection with (3). Also, employing various formulas due to the writer (l. c., 572-4) we obtain a number of congruences each involving only one of the $C(m, i, r)$.

In another paper of the writer's⁴ it was shown that if (1) is satisfied in Case 1 then

$$(6) \quad \sum_{r=1}^{[l/3]} r^{l-3} \equiv 0 \pmod{l},$$

and it was shown by Schwindt⁵ that this yields the relation

$$(6a) \quad \sum_{r=1}^{[l/6]} r^{l-3} \equiv 0 \pmod{l}.$$

From the writer's article last cited (p. 91, and see also last paragraph in article) we also have the criteria

$$(7) \quad \sum' F_2(t/\rho) \frac{f_{l-2}(\rho)}{\rho^l - 1} \equiv 0 \pmod{l}$$

where ρ is an m -th root of unity and \sum' indicates summation over all distinct values $\neq 1$, of $\rho; (m, l) = 1$. Further

$$F_2(w) = w + 2w^2 + \cdots + (ml - 1)w^{ml-1}.$$

The value $m = 3$ gives (6). Set $m = 4$, then we note that

² *Berlin Sitzungsberichte* (1914), pp. 653-81. Cf. also Emma Lehmer, *Annals of Mathematics* (vol. 39 (1938), pp. 358-9.

³ *Japanese Journal of Mathematics*, vol VIII (1931), pp. 159-173.

⁴ *Annals of Mathematics*, vol. 26 (1924), pp. 88-94.

⁵ *Jahresberichte Deutscher Mathematischer Verein*, vol. 43 (1933-4), pp. 229-31.

$$\begin{aligned}\frac{w^{4l}-1}{w-1} &= 1 + w + w^2 + \cdots + w^{4l-1}, \\ w(w^{4l}-1) \frac{d}{dw} \left(\frac{1}{w-1} \right) + \frac{4lw^{4l}}{w-1} &= w + 2w^2 + \cdots + (4l-1)w^{4l-1}, \\ -\frac{w(w^{4l}-1)}{(w-1)^2} &\equiv w + 2w^2 + \cdots + (4l-1)w^{4l-1} \pmod{l}\end{aligned}$$

where we regard a fraction of the form θl , where the denominator is a polynomial in w not all of whose coefficients are divisible by l , as $\equiv 0 \pmod{l}$. Set $w = t/\rho$; we find since $\rho^4 = 1$, $t^l \equiv t \pmod{l}$,

$$-(t/\rho) \frac{t^4-1}{(t/\rho-1)^2} \equiv F_2(t/\rho); \quad -(t/\rho) \frac{(t^4-1)^2}{(t/\rho-1)^2} \equiv (t^4-1)F_2(t/\rho)$$

modulo l , and

$$\sum (-(t/\rho)) \frac{(t^4-1)^2}{(t/\rho-1)^2} \frac{f_{l-2}(\rho)}{\rho^l-1} \equiv (t^4-1) \sum' F_2(t/\rho) \frac{f_{l-2}(\rho)}{\rho^l-1}$$

modulo l . Now

$$\begin{aligned}\frac{t^4-1}{t/\rho-1} &= \frac{(t/\rho)^4-1}{t/\rho-1} = (t/\rho-\rho)(t/\rho-\rho^2)(t/\rho-\rho^3) \\ &= \frac{(t-1)(t-\rho^2)(t-\rho^3)}{\rho^3};\end{aligned}$$

so that (8) becomes, using (7), and noting that $t \not\equiv 0 \pmod{l}$ and $t-1 \not\equiv 0 \pmod{l}$,

$$\sum' \rho(t^2 - (\rho^2 + \rho^3)t + \rho)^2 \frac{f_{l-2}(\rho)}{\rho^l-1} \equiv 0 \pmod{l}$$

or since $\rho^3 + \rho^2 + \rho + 1 = 0$, we have

$$(9) \quad \sum' \rho(t^2 + (\rho+1)t + \rho)^2 \frac{f_{l-2}(\rho)}{\rho^l-1} \equiv 0 \pmod{l}.$$

This congruence is of degree four and since it is satisfied by all the values (2a) then either $t^2 - t + 1 \equiv 0 \pmod{l}$ or $t = -1, 2$ and $1/2$. The first congruence⁶ is inconsistent with (2) and $t = -1$ satisfies (9) identically, but $t = 2$ and $1/2$ give in turn

$$\sum' (4\rho^3 + 4\rho^2 + \rho) \frac{f_{l-2}(\rho)}{\rho^l-1} \equiv 0 \pmod{l},$$

$$\sum' (\rho^3 + 4\rho^2 + 4\rho) \frac{f_{l-2}(\rho)}{\rho^l-1} \equiv 0 \pmod{l},$$

and subtraction gives

$$(9a) \quad \sum' (\rho^3 - \rho) \frac{f_{l-2}(\rho)}{\rho^l-1} \equiv 0 \pmod{l}.$$

⁶ Pollaczek, *Wiener Bericht*, vol. 126 (1917), pp. 1-15.

Now we have ⁷

$$\frac{(4b+3)^{l-2} - b_{l-2}}{l-2} \equiv (-1)^{l-1} \frac{\rho^3 f_{l-2}(\rho)}{\rho^l - 1} \pmod{l},$$

where $(mb+k)^n$ is expanded by the binomial theorem and b_l substituted for b^l in the result. Also

$$\frac{(4b+1)^{l-2} - b_{l-2}}{l-2} \equiv (-1)^{l-1} \sum' \frac{\rho f_{l-2}(\rho)}{\rho^l - 1} \pmod{l},$$

and subtraction of the last two congruences gives

$$(10) \quad \frac{(4b+3)^{l-2} - (4b+1)^{l-2}}{l-2} \equiv \sum' \frac{(\rho^3 - \rho) f_{l-2}(\rho)}{\rho^l - 1}.$$

The left-hand member ⁸ is $(-1)^{(l-3)/2} 2E_{(l-3)/2}$, where $E_1 = 1$, $E_2 = 5$, $E_3 = 61$, etc. are the Euler numbers, and (10) gives with (9a),

$$(11) \quad E_{(l-3)/2} \equiv 0 \pmod{l}.$$

Emma Lehmer ⁹ gave the relation

$$\sum_{r=1}^{[l/4]} \frac{1}{r^2} \equiv \sum_{r=1}^{[l/4]} r^{l-3} \equiv (-1)^{(l-1)/2} 4E_{(l-3)/2}$$

modulo l , and remarked that if we could show the left-hand member $\equiv 0 \pmod{l}$, provided (1) holds in Case 1, that (11) follows. Here we may reverse this process, as using (11) we have

$$\sum_{r=1}^{[l/4]} r^{l-3} \equiv 0 \pmod{l}$$

as criteria for (1) in Case 1, and this is evidently also included in the class of relations (3). Hence we have, using (6a) also,

THEOREM. *If*

$$x^l + y^l + z^l = 0$$

with x, y and z rational integers and $xyz \not\equiv 0 \pmod{l}$; l a given odd prime, then

$$E_{(l-3)/2} \equiv 0 \pmod{l}$$

where $E_1 = 1$, $E_2 = 5$, $E_3 = 61$, . . . , are the Euler numbers. Also

$$\sum_{r=[l/6]+1}^{[l/4]} \frac{1}{r^2} \equiv 0 \pmod{l}.$$

UNIVERSITY OF TEXAS.

⁷ Frobenius, *Sitzungsberichte, Berlin* (1914), p. 655, formula (2) for $n = l - 2$.

⁸ Cf. for example, Frobenius, *Sitzungsberichte, Berlin* (1914), p. 846.

⁹ *Loc. cit.*, p. 359.

ON EXPANSIONS IN SERIES OF EXPONENTIAL FUNCTIONS.*

By MARVIN G. MOORE.

Introduction. Carmichael has expanded functions of exponential type in a series of exponential functions (3) associated with the exponential sum $h(t)$ in (1). He has pointed out that, for the special case $h(t) = e^t - 1$, (3) becomes the Fourier expansion, the natural polygonal region of convergence reducing to the line-segment $(0, 1)$. We are led, then, to investigate the possibility of generalizing the properties of biorthogonality and the convergence theory of the Fourier series to expansions associated with the more general functions $h(t)$.

I. PRELIMINARY CONSIDERATIONS.

Let

$$(1) \quad h(t) = c_1 e^{\alpha_1 t} + c_2 e^{\alpha_2 t} + \cdots + c_N e^{\alpha_N t},$$

where $c_k \neq 0$ and $\alpha_j \neq \alpha_k$ for $j \neq k$, and where $N \geq 2$.

Let P be the smallest closed convex polygon in the complex plane containing the points $\alpha_1, \alpha_2, \dots, \alpha_N$. In special cases, this polygon may reduce to a line-segment. Then Carmichael¹ has demonstrated the existence of contours C_1, C_2, \dots about the origin having the following properties: first, there exists a positive ϵ for which

$$(2) \quad |h(t)e^{-xt}| > \epsilon$$

for every x in P and for every t on every C_s ; second, if the sectors S_μ are defined to be those regions in which $R(\alpha_\mu t) > R(\alpha_\lambda t)$ for all $\lambda \neq \mu$ (R being the real part), C_s lies along the circle having radius s and center at the origin, except for portions of bounded length lying within a bounded distance of the rays which separate the sectors S_μ ; and third, no point of C_s lies outside C_{s+1} .

The series with which we are concerned are to be of the form

$$(3) \quad \sum_{s=1}^{\infty} \sum_{k=1}^{\sigma_s} P_{ks}(x) e^{t_{ks}x},$$

* Presented to the Society, Dec. 30, 1937. Received June 19, 1938; Revised July 1, 1939.

¹ R. D. Carmichael, *Transactions of the American Mathematical Society*, vol. 35, No. 1 (1933), pp. 1-28.

where the degree of the polynomial $P_{ks}(x)$ is at least one less than the order of the zero t_{ks} of $h(t)$, and where $t_{1s}, t_{2s}, \dots, t_{\sigma_s s}$ are those zeroes of $h(t)$ lying between C_{s-1} and C_s .

II. PROPERTIES OF BIORTHOGONALITY.

THEOREM 1. *Let $w(k, s)$ be the order of the zero t_{ks} of $h(t)$. Let C_{ks} be a small circle passing through no zero of $h(t)$ and containing only the zero t_{ks} on its interior. Then, for $q = 0, 1, 2, \dots, w(k, s) - 1$, and for a any point of the closed region P ,*

$$(4) \quad \frac{1}{2\pi i} \sum_{\mu=1}^N c_{\mu} \int_a^{a_{\mu}} x_1^q e^{t_{ks} x_1} \int_{C_{jm}} \frac{e^{(a_{\mu} + x - x_1)t}}{h(t)} dt dx_1$$

$$\left\{ \begin{array}{l} = 0 \text{ for } t_{ks} \neq t_{jm} \\ = x^q e^{t_{ks} x} \text{ for } t_{ks} = t_{jm}. \end{array} \right.$$

Upon integration with respect to x_1 , the left member of (4) reduces to the form

$$- \frac{1}{2\pi i} \int_{C_{jm}} (-1)^{\nu} \frac{a^{q-\nu} q! e^{(t_{ks}-t)a+xt}}{(q-\nu)! (t_{ks}-t)^{\nu+1}} dt,$$

which, by Cauchy's Integral Theorem, vanishes if $t_{ks} \neq t_{jm}$.

If $t_{ks} = t_{jm}$, it reduces to

$$\sum_{\nu=0}^q (x-a)^{\nu} a^{q-\nu} \binom{q}{\nu} e^{t_{ks} x},$$

where $\binom{q}{\nu}$ is the binomial coefficient, and the expression, by the binomial theorem, equals

$$x^q e^{t_{ks} x}.$$

We have, then, conditions generalizing the biorthogonality conditions pertaining to the Fourier series, here arising when $h(t) = e^t - 1$, our results in that case taking the form, after the contour integrals are evaluated: If a is any fixed point of the interval $(0, 1)$, then for m and l integers, positive, negative, or zero,

$$e^{2l\pi i(x+1)} \int_a^1 e^{2(m-l)\pi i x_1} dx_1 - e^{2l\pi i x} \int_a^0 e^{2(m-l)\pi i x_1} dx_1$$

$$\left\{ \begin{array}{l} = 0 \text{ for } m \neq l \\ = e^{2m\pi i x} \text{ for } m = l. \end{array} \right.$$

If $a = 0$, this reduces essentially to the customary form of the statement of the Fourier biorthogonality conditions.

Theorem 1, viewed in the light of the theory of biorthogonal functions, suggests the examination of the series (see (3))

$$(5) \quad \sum_{s=1}^{\infty} \sum_{k=1}^{\sigma_s} \frac{1}{2\pi i} \sum_{\mu=1}^N c_{\mu} \int_a^{a_{\mu}} f(x_1) \int_{C_{ks}} e^{(a_{\mu} x - x_1)t} \{h(t)\}^{-1} dt dx_1,$$

which we shall call *the F-series*. Series (5) is of the form (3).

III. LEMMAS ON CONTOUR INTEGRALS.

We shall find it convenient to define P_{η} as the region P exclusive of the portions $|x - \alpha_{\mu}| < \eta$ about the vertices.

LEMMA 1. *For every positive η , there exists a positive K for which*

$$\left| \int_{C_s} e^{xt} \{h(t)\}^{-1} dt \right| < K$$

for $s = 1, 2, \dots$, and for x in P_{η} .

Since, by (2), the integrand is dominated in absolute value by ϵ^{-1} at every point t of C_s , for all s , and for all x in P , it follows that the portion of C_s which does not lie on the circle of radius s , having for its length a bounded function of s , contributes a bounded quantity to the value of the above integral.

To show that a bounded quantity is also contributed by each of the circular arcs of C_s which remain; for any sector S_{λ} , let $x - \alpha_{\lambda} = re^{i\psi}$, $t = \rho e^{i\phi}$; where ψ, ϕ are real and r, ρ are positive. It follows almost immediately from the definition of S_{λ} that, for x in P and for t in S_{λ} , $R(xt) \leq R(\alpha_{\lambda}t)$, so that

$$\cos(\psi + \phi) \leq 0.$$

On the arc of C_s in S_{λ} , as well as at all other points of C_s and for all s , $|e^{\alpha_{\lambda}t} \{h(t)\}^{-1}| < \epsilon^{-1}$, so that we shall, along this arc, dominate the integral by

$$\epsilon^{-1} \int |e^{(x-\alpha_{\lambda})t}| |dt| \leq \epsilon^{-1} \int_{(1/2)\pi}^{(3/2)\pi} e^{r\rho \cos x} \rho dx \quad (x = \psi + \phi).$$

Since, for $\frac{1}{2}\pi \leq x \leq \pi$, $\cos x \leq -2\pi^{-1}(x - \frac{1}{2}\pi)$, we may dominate this expression by

$$2\epsilon^{-1} \int_{(1/2)\pi}^{\pi} e^{-2r\rho\pi^{-1}(x-\frac{1}{2}\pi)} \rho dx = \pi r^{-1} \epsilon^{-1} (1 - e^{-r\rho}) < \pi \eta^{-1} \epsilon^{-1}$$

for $r \geq \eta$. The lemma has then been proved.

LEMMA 2. Let θ_μ be the supplement of the angle of P at α_μ if α_μ is a vertex of P , and let it otherwise be zero. Then

$$\lim_{s \rightarrow \infty} \int_{C_s} \frac{e^{a_\mu t}}{h(t)} \frac{dt}{t} = \frac{i\theta_\mu}{c_\mu}.$$

We shall consider only the case for which α_μ is a vertex, for otherwise the conclusion is a special case of Carmichael's result that

$$(6) \quad \lim_{s \rightarrow \infty} \int_{C_s} e^{xt} \{h(t)\}^{-1} t^{-1} dt = 0$$

for x in P and not a vertex.²

Breaking C_s up into C' and C'' , where C' is that portion of C_s in S_μ ,

$$c_\mu \int_{C_s} \frac{e^{a_\mu t}}{h(t)} \frac{dt}{t} - i\theta_\mu = \int_{C'} \left[\frac{c_\mu e^{a_\mu t}}{h(t)} - 1 \right] \frac{dt}{t} + \int_{C''} \frac{c_\mu e^{a_\mu t}}{h(t)} \frac{dt}{t}.$$

Take parabolas³ with vertices at the origin and with the bounding rays of S_μ as principal diameters. Writing $t = \rho e^{i\phi}$, we then see that the resulting integrands approach zero for t outside the parabolas, for s becoming infinite, and are bounded for t inside the parabolas. The ranges of integration with respect to ϕ approach zero inside the parabolas, so that the integrals with respect to ϕ approach zero. The integrals with respect to ρ likewise approach zero, the integrands approaching zero in that case.

IV. CONVERGENCE OF THE F-SERIES.

Let P' be a convex polygon contained in P and let it have the property that for every point α_μ ($\mu = 1, 2, \dots, N$), there exists a point α'_μ in P' for which $\alpha_\mu + x - \alpha'_\mu$ lies in P and not at a vertex of P , for all x in P' and not vertices of P' . In particular, P' may coincide with P , in which case $\alpha'_\mu = \alpha_\mu$.

Let a curve H_μ from each point α'_μ to the corresponding point α_μ be made up of a finite number of straight-line segments and let it have the property that, for every point x_1 on H_μ and for every x in P' and not a vertex of P' , $\alpha_\mu + x - x_1$ lies in P and not at a vertex of P . In particular, the curves H_μ may be taken to be straight lines, although, if straight lines are not suitable, we are not, in general, restricted to them.

² Carmichael, *loc. cit.*, p. 24.

³ Carmichael uses such parabolas for similar purposes, *loc. cit.*, p. 24.

We shall find it convenient to use the notation

$$f[x + 0(\alpha'_\mu - x)] = \lim_{\omega} f[x + \omega(\alpha'_\mu - x)] (\omega \rightarrow 0; \omega > 0),$$

when the limit exists, in the following theorem.

Let $f(x_1)$ be so defined that both its real and imaginary parts are summable (L) along each of the line segments of which the curves H_μ are formed and also along every straight line segment in closed P' . Further, if P' does not reduce to a straight line segment, let $f(x_1)$ be analytic in open P' and let its integral between any two points of closed P' (taken along any finite number of straight-line segments) be independent of the path of integration in P' . Let a be any point of P' and let the paths of integration L_μ from a to α_μ ($\mu = 1, 2, \dots, N$) be made up of the straight lines from a to α'_μ combined with the curves H_μ .

THEOREM 2. *Let the above hypotheses be satisfied.*

Then, first, for every point x on the interior of P' (if there be such points); and second, for every x on the boundary, and not at a vertex of P' , for which there exists a positive number η , such that both the real and imaginary parts of $f(x_1)$ are of bounded variation in the linear interval $[x, x + \eta_1(\alpha'_\mu - x)]$ ($\mu = 1, 2, \dots, N$); the F -series for $f(x)$ associated with $h(t)$ and L_μ ($\mu = 1, 2, \dots, N$) converges to

$$\frac{1}{2} \sum_{\mu=1}^N \theta_\mu \pi^{-1} f[x + 0(\alpha'_\mu - x)].$$

For x on the interior of P' , this expression is equal to $f(x)$.

On breaking up the functions $f(x_1)$ and

$$(7) \quad \int_{C_k} e^{(\alpha_\mu + x - x_1)t} \{h(t)\}^{-1} dt$$

(which we shall write as Q) into their real and imaginary parts for x_1 on any straight-line segment (β, γ) on L_μ , since both the real and imaginary parts of Q have continuous derivatives with respect to x_1 ,

$$(8) \quad \int_{\beta}^{\gamma} f(x_1) Q dx_1$$

(which we shall write as $I(\beta, \gamma)$) may be written as the sum of real and imaginary Lebesgue integrals, each of which may be integrated by parts,⁴ so that

⁴ E. W. Hobson, *The Theory of Functions of a Real Variable*, vol. I (1927), p. 216.

$$(9) \quad I(\beta, \gamma) = \int_{\beta}^{\gamma} f dx_1 \int_{\beta}^{\gamma} \frac{dQ}{dx_1} dx_1 - \int_{\beta}^{\gamma} \frac{d}{dx_1} Q \int_{\beta}^{x_1} f(x_2) dx_2 dx_1$$

If (β, γ) lies in P' , both factors of the first term of (9) are independent of the path of integration in P' , while both factors of the integrand in the second term are continuous throughout closed P' and analytic on the interior, so that the second term of (9) is independent of the path of integration in P' , and (8) must be also.

Now let a_1 and a_2 be any two choices of the point a . Then, under our hypotheses,

$$\sum_{\mu} c_{\mu} \{I(a_1, \alpha_{\mu}) - I(a_2, \alpha_{\mu})\} = \int_{a_1}^{a_2} f(x_1) \int_{C_{k\mu}} e^{(x-x_1)t} dt dx_1,$$

which, by the Cauchy Integral Theorem, vanishes, so that every term of (5) is independent of a , for a in P' .

We may then (and shall) take a at the point x , writing, by the theory of residues, the sum of the first s terms of (5) in the form

$$\frac{1}{2\pi i} \sum_{\mu} c_{\mu} J(x, \alpha_{\mu}),$$

where U and J replace Q and I respectively for $C_{k\mu}$ replaced by C_{μ} . We shall then consider the separate terms (for convenience dropping the subscript μ), which may be written as

$$(10) \quad \frac{c}{2\pi i} \int_x^a f[x + 0(\alpha' - x)] U dx_1 \\ + \frac{c}{2\pi i} \int_x^a \{f(x_1) - f[x + 0(\alpha' - x)]\} U dx_1.$$

Upon evaluation of the first integral in (10), its limit as s becomes infinite is seen, by (6) and lemma 2, to be equal to

$$\frac{1}{2\pi} \theta_{\mu} f[x + 0(\alpha'_{\mu} - x)].$$

It will, then, be sufficient for our proof to show that the limit of the second term of (10) vanishes.

We now set

$$R\{f(x_1) - f[x + 0(\alpha' - x)]\} = A_1(x, x_1) - A_2(x, x_1)$$

for x_1 on L (where L now joins x to α), A_1 and A_2 being monotonic functions of x_1 for x_1 in the open interval $[x, x + \eta_1(\alpha' - x)]$, approaching zero as $x_1 \rightarrow x$ and being summable on L . Then for every positive ζ there exists a

positive $\eta < \eta_1$ for which $|A_1(x, x_1)| < \xi$ for x_1 in the linear interval $[x, x + \eta(\alpha' - x)]$, so that, upon using the second mean value theorem,⁵ we see that

$$\int_{x_1=x}^{x_1=x+\eta(\alpha'-x)} A_1(x, x_1) R(U) dR(x_1)$$

may be dominated in absolute value by

$$(11) \quad \xi \left| \int_{x_1=\xi_s}^{x_1=x+\eta(\alpha'-x)} R(U) dR(x_1) \right|,$$

where ξ_s lies in the interval $[x, x + \eta(\alpha' - x)]$. Since $dR(x_1)/dx_1$ is to be constant, we may actually carry out the integration, and then, by (2), we find that (11) is dominated by $4\pi\xi\epsilon^{-1}$.

Take now any straight line portion (β, γ) of the part of L not involved in (11) and let us consider

$$(12) \quad \int_{x_1=\beta}^{x_1=\gamma} A_1(x, x_1) R(U) dR(x_1).$$

As we prepare to apply Hobson's General Convergence Theorem,⁶ we note first that, by our hypotheses, $\alpha + x - x_1$ lies in P and not at a vertex for every x_1 on closed (β, γ) . It must, then, be bounded away from the vertices, so that for every positive η there exists an $\bar{\eta}$ for which $\alpha + x - x_1$ lies in $P_{\bar{\eta}}$. By lemma 1, there then exists a positive K for which $|U| < K$ for all s , so that $|R(U)|$ is also bounded, and Hobson's first condition is satisfied.

Noting that

$$\left| \int_{x_1=\beta}^{x_1=\gamma} R(U) dR(x_1) \right| \leq \left| \int_{x_1=\beta}^{x_1=\gamma} U dR(x_1) \right| = \left| \frac{R(\gamma) - R(\beta)}{\gamma - \beta} \right| \int_{\beta}^{\gamma} U dx_1,$$

we carry out the integration of U with respect to x , and then apply (6) to show that the second condition is also satisfied.

Then, for every positive η , (12) approaches zero as s becomes infinite.

We may then combine the finite number of line-segments which form L to obtain the result: For every positive η and for every positive ξ , there exists a positive integer \bar{s}_1 for which

$$(13) \quad \left| \int_{x_1=x}^{x_1=\alpha} A_1(x, x_1) U dR(x_1) \right| < \xi + 4\pi\xi\epsilon^{-1}$$

for $s > \bar{s}_1$.

⁵ E. W. Hobson, *loc. cit.*, p. 618.

⁶ Reference will be made to E. W. Hobson, *The Theory of Functions of a Real Variable*, vol. II (1926), p. 422; see also *Proceedings of the London Mathematical Society* (2), vol. VI (1908), p. 349, and (2), vol. XII (1912), p. 166.

We now note that the second term in (10) may, by the separation of the real and imaginary parts of the factors of the integrand, be divided up into a finite number of parts each of which may be given essentially the same treatment as we have given the integral in (13), so that the second term in (10) approaches zero as s becomes infinite, and the F-series converges to

$$\frac{1}{2} \sum_{\mu} \theta_{\mu} \pi^{-1} f[x + 0(\alpha'_{\mu} - x)].$$

In particular, if $h(t) = e^t - 1$ so that we are dealing with the Fourier series on the interval $(0, 1)$, we find the series to converge to

$$\frac{1}{2}\{f(x+0) + f(x-0)\}.$$

It may be shown by similar methods, for P' coinciding with P , that, at the vertex α_{λ} , the F-series converges to

$$\frac{1}{2} \sum_{\mu \neq \lambda} \theta_{\mu} \pi^{-1} f[\alpha_{\lambda} + 0(\alpha_{\mu} - \alpha_{\lambda})] - \frac{1}{2} \sum_{\mu \neq \lambda} \theta_{\lambda} \pi^{-1} c_{\mu}/c_{\lambda} f[\alpha_{\mu} + 0(\alpha_{\lambda} - \alpha_{\mu})].$$

In particular, the Fourier series converges to

$$\frac{1}{2}\{f(1-0) + f(+0)\}$$

at either end-point of the interval.

INDIANA UNIVERSITY,
BLOOMINGTON, INDIANA.

EXTREMAL PROBLEMS FOR FUNCTIONS ANALYTIC AND SINGLE-VALUED IN A DOUBLY-CONNECTED REGION.*

By MAURICE H. HEINS.

1. Introduction. It is well-known that certain fundamental inequalities of analysis such as Julia's Principle of the Harmonic Majorant,¹ the Two Constant Theorem,² Lindelöf's Principle,³ the Principle of Hyperbolic Measure⁴ are the "best possible" when the domain of definition G_z for the functions $w(z)$ involved is simply-connected. When one considers, however, functions which are analytic and *uniform* in a multiply-connected region, these inequalities are, in general, no longer the "best possible"; it is then a question of interest to determine effectively exact bounds and the associated extremal functions for these inequalities when we restrict our attention to functions which are analytic and *single-valued* in a given multiply-connected region G_z .

By application of the Poincaré Uniformisation Theorem⁵ and the Pick-Nevalinna theory of interpolation⁶ one can determine effectively the exact bounds and the associated extremal functions for the inequalities cited above for the case where G_z is *doubly-connected* and has as its boundary two disjoint continua. To this end we shall study the problem of interpolation for bounded functions which are analytic in the unit circle and satisfy a given functional relation. By the results of this study we shall give a method for determining effectively the exact bounds at a given point z_0 of G_z and the associated extremal functions for the following inequalities: 1) Julia's Principle of the Harmonic Majorant of which the Nevanlinna-Ostrowski Two Constant Theorem and Hadamard's Three Circle Theorem are special cases, 2) the Principle of Hyperbolic Measure of which the Aumann-Carathéodory "Starr-

* Received March 6, 1939.

¹ G. Julia, *Principes géométriques d'analyse*, 2ième partie (Paris, 1932), pp. 26-27.

² R. Nevanlinna, *Eindeutige Analytische Funktionen* (Berlin, 1936), pp. 41-42.

³ E. Lindelöf, "Mémoire sur certaines inégalités dans la théorie des fonctions monogènes et sur quelques propriétés nouvelles de ces fonctions dans le voisinage d'un point singulier essentiel," *Acta. Soc. Sci. Fenn.*, 35 Nr. 7 (1908).

⁴ R. Nevanlinna, *l. c.*, pp. 45-51.

⁵ H. Poincaré, "Sur l'uniformisation des fonctions analytiques," *Acta Mathematica*, vol. 31 (1907).

⁶ R. Nevanlinna, "Ueber beschränkte analytische Funktionen," *Ann. Acad. Sci. Fenn.*, vol. 32, No. 7.

heitssatz" ⁷ is a special case. In addition, our preliminary study permits the complete treatment of the analogue of the Pick-Nevanlinna problem for the case where the interpolating functions are analytic and single-valued in a doubly-connected region. Lammell has considered related interpolation problems.

Recently Carlson ⁸ and Teichmüller ⁹ have considered the problem of improving the Hadamard Three Circle Theorem for functions which are uniform. Their methods are quite distinct from ours, which admit application to other problems as well—the extremal problem for the Principle of Hyperbolic Measure and the Pick-Nevanlinna interpolation problem for doubly-connected regions.

The author wishes to express his thanks to Professor Walsh for his helpful discussions during the preparation of this paper.

2. The Pick-Nevanlinna theory of interpolation. In this section we shall state briefly the principal results of the Pick-Nevanlinna theory of interpolation important for the sequel. For a detailed account of this theory the reader is referred to the treatise of Walsh.¹⁰

Let \mathcal{E} denote the class of functions $w(z)$ analytic for $|z| < 1$ and satisfying there the inequality $|w(z)| \leq 1$. Let α be any complex number for which $|\alpha| < 1$. We denote by $L(z, \alpha)$ the linear fractional function

$$\frac{\alpha - z}{1 - \bar{\alpha}z}.$$

Further let the points z_1, \dots, z_n be given interior to the unit circle $|z| = 1$ and let there be associated with each z_k a complex number $w_k^{(0)}$, $|w_k^{(0)}| \leq 1$ ($k = 1, 2, \dots, n$). Define $w_2^{(1)}, \dots, w_n^{(1)}$ by

$$(2.1) \quad L(w_k^{(0)}, w_1^{(0)}) \equiv \frac{|z_1|}{z_1} L(z_k, z_1) L(w_k^{(1)}, |z_1| w_1^{(0)}) \quad (k = 2, \dots, n)$$

(where $\frac{|z_1|}{z_1} L(z, z_1)$ is to be replaced by z , if $z_1 = 0$), and, in general, $w_k^{(v)}$ ($k = v + 1, \dots, n$) by the recursive formula

⁷ G. Aumann and C. Carathéodory, "Ein Satz über die konforme Abbildung mehrfach-zusammenhängende ebene Gebiete," *Mathematische Annalen*, vol. 109, pp. 756-763.

⁸ F. Carlson, "Sur le module maximum d'une fonction analytique uniforme," *Ark. för Mat. Astron. och Fys.* Bd. 26, 2A9, pp. 1-13.

⁹ O. Teichmüller, "Eine Verschärfung des Dreikreisesatzes," *Deutsche Mathematik* vol. 1 (1939), pp. 16-22.

¹⁰ J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, New York, pp. 286-304.

$$(2.2) \quad L(w_k^{(\nu-1)}, w_\nu^{(\nu-1)}) \equiv \frac{|z_\nu|}{z_\nu} L(z_k, z_\nu) L(w_k^{(\nu)}, |z_\nu| w_\nu^{(\nu-1)})$$

$$(k = \nu + 1, \dots, n)$$

(where $\frac{|z_\nu|}{z_\nu} L(z, z_\nu)$ is to be replaced by z , if $z_\nu = 0$). Then we have

THEOREM 2.1. *A necessary and sufficient condition that there exist a function $w(z) \in \mathcal{E}$ for which $w(z_k) = w_k^{(0)}$ where the z_k are n distinct points interior to $|z| = 1$ is that either*

$$1) \quad |w_1^{(0)}| < 1, |w_2^{(1)}| < 1, \dots, |w_\mu^{(\mu-1)}| < 1, |w_{\mu+1}^{(\mu)}| = 1,$$

$$w_{\mu+1}^{(\mu)} = w_{\mu+2}^{(\mu)} = \dots = w_n^{(\mu)}.$$

$$\text{or } 2) \quad |w_1^{(0)}| < 1, |w_2^{(1)}| < 1, \dots, |w_n^{(n-1)}| < 1.$$

If 1) occurs, $w(z)$ which satisfies the interpolation requirements is unique and is given by the formulas (2.1) and (2.2) in conjunction with

$$(2.3) \quad L(w_0(z), w_1^{(0)}) \equiv \frac{|z_1|}{z_1} L(z, z_1) L(w_1(z), |z_1| w_1^{(0)})$$

$$(2.4) \quad L(w_{\nu-1}(z), w_\nu^{(\nu-1)}) \equiv \frac{|z_\nu|}{z_\nu} L(z, z_\nu) L(w_\nu(z), |z_\nu| w_\nu^{(\nu-1)})$$

$$(2.5) \quad w_\nu(z_k) = w_k^{(\nu)} \quad (k = \nu + 1, \dots, n),$$

where $w_0(z) \equiv w(z)$.

If 2) occurs, $w(z)$ is not unique. All such functions and only such functions are given by the formulas (2.2), (2.4), and (2.5), where $w_n(z)$ is any function of class \mathcal{E} .

Further, if z_1, z_2, \dots ($|z_k| < 1$) are infinite in number, Theorem 2.1 admits the extension

THEOREM 2.2. *A necessary and sufficient condition that there exist a function $w(z) \in \mathcal{E}$ for which $w(z_k) = w_k^{(0)}$ ($k = 1, 2, \dots$) is that either*

$$1) \quad |w_1^{(0)}| < 1, |w_2^{(1)}| < 1, \dots, |w_\mu^{(\mu-1)}| < 1, |w_{\mu+1}^{(\mu)}| = 1$$

$$w_{\mu+1}^{(\mu)} = w_{\mu+2}^{(\mu)} = \dots$$

$$\text{or } 2) \quad |w_1^{(0)}| < 1, |w_2^{(1)}| < 1, \dots$$

If 1) occurs, $w(z)$ with the required properties is unique and is given by the recursive formulas (2.2), (2.4), and (2.5) where $w_\mu(z) \equiv w_{\mu+1}^{(\mu)}$.

In the situation of Theorem 2.2 we have

THEOREM 2.3. *If there exists a function $w(z) \in \mathcal{E}$ satisfying the interpolation requirements of Theorem 2.2, then all $w(z) \in \mathcal{E}$ satisfying these requirements can be expressed in the form*

$$(2.6) \quad w(z) = \frac{P(z) - Q(z)w_{\infty}(z)}{1 - S(z)w_{\infty}(z)}$$

where P , Q , and S are specific functions of class \mathcal{E} defined by the interpolation requirements and $w_{\infty}(z)$ is an arbitrary function of class \mathcal{E} . And conversely, every function $w(z)$ defined by (2.6) where $w_{\infty}(z)$ is an arbitrary function of class \mathcal{E} belongs to class \mathcal{E} and satisfies the interpolation requirements of Theorem 2.2.

A necessary and sufficient condition that $w \in \mathcal{E}$ satisfying the interpolation requirements of Theorem 2.2 be unique is that

$$PS - Q \equiv 0.$$

Remark. If $PS - Q \not\equiv 0$, the function $PS - Q$ vanishes at the points z_k and at no other points for $|z| < 1$.

3. A particular interpolation problem. We now turn our attention to a particular interpolation problem which is fundamental for the study that we shall make. Let $Tz (\not\equiv z)$ denote any linear fractional transformation mapping $|z| \leq 1$ onto itself (in the sequel we shall consider exclusively the case where T is hyperbolic), and let Uz denote a second such transformation, but here we do not require that $Uz \not\equiv z$; in fact, the case where $Uz \equiv z$ is of prime importance. We wish to study those functions $w(z) \in \mathcal{E}$ which satisfy certain interpolation requirements at assigned points z_k ($k = 1, 2, \dots, n$ or $k = 1, 2, \dots$) and which satisfy for $|z| < 1$ the functional relation

$$(3.1) \quad \boxed{w(T) = U[w(z)]}.$$

We shall demonstrate the following

THEOREM 3.1. *A necessary and sufficient condition that there exist a function $w(z) \in \mathcal{E}$ for which*

$$w(z_k) = w_k^{(0)}, \quad |z_k| < 1, \quad w(T^m z_k) = U^m w_k^{(0)} \\ (k = 1, 2, \dots, n \text{ or } k = 1, 2, \dots; m = 0, \pm 1, \pm 2, \dots)$$

(it is assumed that these interpolation requirements are consistent) and which satisfies the functional relation (3.1) is that there exist a function $w^*(z) \in \mathcal{E}$ which satisfies the interpolation requirements for $w(z)$.

It is clear that this condition is necessary.

To prove that it is sufficient we note that if $w^*(z)$, the existence of which is posited, is unique, then

$$w^*(T) = U[w^*(z)],$$

for $U^{-1}[w^*(T)] \subset \mathcal{E}$ and satisfies the same interpolation requirements as $w^*(z)$. Therefore

$$U^{-1}[w^*(T)] = w^*(z)$$

or

$$w^*(T) = U[w^*(z)].$$

If $w^*(z)$ is not unique, let $\{w^*(z)\}$ denote the totality of functions $w^*(z)$ which satisfy the interpolation requirements and let z_0 be any point distinct from all the points $T^m z_k$

$$(k = 1, 2, \dots, n \text{ or } k = 1, 2, \dots; m = 0, \pm 1, \pm 2, \dots).$$

Then $\{w^*(z_0)\}$ is the totality of values which the functions $w^*(z) \subset \{w^*(z)\}$ take on at z_0 . Consider the set $\{U^{-1}[w^*(Tz_0)]\}$. We assert that

$$(3.2) \quad \{w^*(z_0)\} = \{U^{-1}[w^*(Tz_0)]\}.$$

Let $w^*_1(z) \subset \{w^*(z)\}$, then it follows that $U^{-1}[w^*_1(T)] \subset \mathcal{E}$ and satisfies the interpolation requirements; therefore $U^{-1}[w^*_1(Tz_0)] \subset \{w^*(z_0)\}$ and therefore

$$\{U^{-1}[w^*(Tz_0)]\} \subset \{w^*(z_0)\}.$$

Also if $w^*_2(z) \subset \{w^*(z)\}$, then $w^*_2(z) = U^{-1}[U[w^*_2(T^{-1}Tz)]]$. But $U[w^*_2(T^{-1})] \subset \mathcal{E}$ and satisfies the interpolation requirements. Therefore $w^*_2(z_0) \subset \{U^{-1}[w^*(Tz_0)]\}$ and it follows that

$$\{w^*(z_0)\} \subset \{U^{-1}[w^*(Tz_0)]\}.$$

Therefore the relation (3.2) is verified. From this fact we shall deduce certain functional relations between $P(z)$, $Q(z)$, $S(z)$ and $P(T)$, $Q(T)$, $S(T)$. Let us note that since the solution of the interpolation problem is not unique and since z_0 is distinct from $T^m z_k$ ($k = 1, 2, \dots, n$ or $k = 1, 2, \dots$; $m = 0, \pm 1, \pm 2, \dots$), $P(z_0)Q(z_0) - S(z_0) \neq 0$ (cf. remark Theorem 2.3), and therefore the set $\{w^*(z_0)\}$ fills a proper circle which we shall denote by K_{z_0} . This follows from the formula (2.6) where $w(z)$ is replaced by $w^*(z)$ and the statement of the interpolation requirements of Theorem 2.3 is replaced by the statement of the interpolation requirements of the theorem which we are to prove. The transformation

$$(3.3) \quad w^* = \frac{P(z_0) - Q(z_0)w^*_\infty}{1 - S(z_0)w^*_\infty}$$

maps $|w^*_\infty| \leq 1$ onto K_{z_0} ; and the transformation

$$(3.4) \quad w^* = U^{-1} \left[\frac{P(Tz_0) - Q(Tz_0)w^*_\infty}{1 - S(Tz_0)w^*_\infty} \right]$$

maps $|w^*_\infty| \leq 1$ onto K_{z_0} by virtue of the relation (3.2). Therefore the transformation ¹¹

$$(3.5) \quad \frac{P(z_0) - Q(z_0)t}{1 - S(z_0)t} = U^{-1} \left[\frac{P(Tz_0) - Q(Tz_0)\tau}{1 - S(Tz_0)\tau} \right]$$

is non-degenerate and maps $|t| \leq 1$ onto $|\tau| \leq 1$. Let us write $U(z)$ in the form

$$e^{i\theta}(z - \alpha)/(\bar{\alpha}z - 1) \quad (|\alpha| < 1, 0 \leq \theta < 2\pi).$$

Then (3.5) takes the equivalent form

$$(3.6) \quad \frac{P(Tz_0) - Q(Tz_0)\tau}{1 - S(Tz_0)\tau} = \frac{e^{i\theta} \left[\frac{P(z_0) - \alpha}{\bar{\alpha}P(z_0) - 1} - \frac{Q(z_0) - \alpha S(z_0)}{\bar{\alpha}P(z_0) - 1} t \right]}{1 - \frac{\bar{\alpha}Q(z_0) - S(z_0)}{\bar{\alpha}P(z_0) - 1} t}.$$

It follows that (3.6) can be written in the form

$$\tau = [\lambda(z_0) + \epsilon(z_0)t]/[1 + \bar{\lambda}(z_0)\epsilon(z_0)t]$$

where $\lambda(z_0)$ and $\epsilon(z_0)$ are suitably chosen, $|\lambda(z_0)| < 1$, $|\epsilon(z_0)| = 1$. A necessary and sufficient condition that the transformation (3.6) can be written in this form is that the following equations be satisfied:

$$(3.7) \quad \begin{aligned} \frac{e^{i\theta}(P(z_0) - \alpha)}{\bar{\alpha}P(z_0) - 1} &= \frac{P(Tz_0) - \lambda(z_0)Q(Tz_0)}{1 - \lambda(z_0)S(Tz_0)}, \\ \frac{e^{i\theta}(Q(z_0) - \alpha S(z_0))}{\bar{\alpha}P(z_0) - 1} &= \epsilon(z_0) \frac{Q(Tz_0) - \bar{\lambda}(z_0)P(Tz_0)}{1 - \lambda(z_0)S(Tz_0)}, \\ \frac{\bar{\alpha}Q(z_0) - S(z_0)}{\bar{\alpha}P(z_0) - 1} &= \epsilon(z_0) \frac{S(Tz_0) - \bar{\lambda}(z_0)}{1 - \lambda(z_0)S(Tz_0)}. \end{aligned}$$

Now the equations (3.7) with the subscript dropped determine $\epsilon(z)$, $\lambda(z)$, $\bar{\lambda}(z)$ as functions of z , single-valued, analytic and defined for all values of z interior to the unit circle $|z| = 1$ other than the $\{T^m z_k\}$. The conditions $|\epsilon(z)| = 1$ and $\lambda(z)\bar{\lambda}(z)$ real for every such z imply that ϵ and λ are con-

¹¹ *Ibid.* In particular, pp. 296-304.

stant. Therefore the relations (3.7) are valid for all z , $|z| < 1$ where ϵ , λ are constant.

Suppose now that $w(z)$ satisfies the required interpolation conditions and further the functional relation (3.1). Then $w(z)$ can be written in the form

$$(3.8) \quad w(z) = \frac{P(z) - Q(z)w_{\infty}(z)}{1 - S(z)w_{\infty}(z)},$$

and $w_{\infty}(z)$ satisfies the functional relation

$$(3.9) \quad \frac{P(T) - Q(T)w_{\infty}(T)}{1 - S(T)w_{\infty}(T)} = U \left[\frac{P(z) - Q(z)w_{\infty}(z)}{1 - S(z)w_{\infty}(z)} \right],$$

and from our discussion we infer that

$$(3.10) \quad w_{\infty}(T) = U^*[w_{\infty}(z)],$$

where

$$U^*(\tau) = (\lambda + \epsilon\tau)/(1 + \bar{\lambda}\epsilon\tau).$$

Conversely, if $w_{\infty}(z) \in \mathcal{E}$ and satisfies the functional relation (3.10), $w(z)$ as given by (3.8) satisfies the required interpolation conditions and further the functional relation (3.1). Now there always exists a function $w_{\infty}(z) \in \mathcal{E}$ which satisfies the functional relation (3.10). For since U^* maps the closed interior of the unit circle onto itself, it is either the identical transformation or a transformation of one of the following types: elliptic, hyperbolic, parabolic. If U^* is the identical transformation, the existence of a function $w_{\infty}(z) \in \mathcal{E}$ satisfying (3.10) is evident; any constant k , $|k| \leq 1$ satisfies (3.10). If U^* is not the identical transformation, then it is well-known from the theory of linear fractional transformations that $U^*(\tau)$ has at least one fixed point in the closed interior of the unit circle $|\tau| \leq 1$. Let τ^* be a fixed point of $U^*(\tau)$; then it is evident that τ^* satisfies the relation (3.10). (We shall return to the study of the functional equation (3.10) and consider the possibility of non-constant solutions.) Thus Theorem 3.1 is established.

Let us remark that when $w^*(z)$ satisfying the interpolation requirements is not unique, there is a one-to-one correspondence between the functions $w_{\infty}(z)$ satisfying the functional relation (3.10) and the functions $w(z)$ which satisfy the interpolation conditions of Theorem 3.1 and the functional equation (3.1).

Denjoy has given the following criterion for the uniqueness of $w(z)$ of

Theorem 2.2: ¹² A necessary and sufficient condition that the function $w(z)$ of Theorem 2.2 2) be unique is the divergence of the series

$$(3.11) \quad \sum_{\nu=1}^{\infty} \frac{1 - |z_{\nu}|}{1 - |w_{\nu}^{(\nu-1)}|}.$$

Let us remark that in the applications which we shall make of Theorem 3.1, we shall consider exclusively the case where T is hyperbolic. We shall demonstrate that if T is hyperbolic, a necessary and sufficient condition that $w(z)$ of Theorem 3.1 be unique is that the function $w^*(z)$ of Theorem 3.1 be unique. But a necessary and sufficient condition for the uniqueness of $w^*(z)$ and therefore for the uniqueness of $w(z)$ is the criterion of Denjoy (3.11) where the notation is suitably modified.

As we have shown, if $w^*(z)$ of Theorem 3.1 is unique, then $w(z)$ is also unique. Suppose now that $w^*(z)$ is not unique. We shall show that $w(z)$ is not unique.

Let us recall that there is a one-to-one correspondence between the $w(z)$ of Theorem 3.1 and the functions $w_{\infty}(z) \subset \mathcal{E}$ which satisfy the functional relation (3.10). If U^* of the relation (3.10) is the identical transformation or hyperbolic, there is more than one solution of (3.10) which belongs to class \mathcal{E} . Two cases remain. U^* may be parabolic or elliptic. But in these cases the relation (3.10) may be reduced to the following canonical forms:

A) U^* elliptic

$$\begin{aligned} f(\lambda z) &= e^{i\theta} f(z), & R(z) &> 0, & |f| &\leq 1 \\ \theta \text{ real}, & \lambda \text{ positive} & (\neq 1), \end{aligned}$$

B) U^* parabolic

$$\begin{aligned} f(\lambda z) &= f(z) + i\mu, & R(z) &> 0, & R(f) &\geq 0 \\ \mu \text{ real}, & \lambda \text{ positive} & (\neq 1). \end{aligned}$$

Let us consider Case A). It is clear that the function

$$K e^{i\theta \log z / \log \lambda}$$

satisfies the equation

$$f(\lambda z) = e^{i\theta} f(z)$$

where K is a constant. We seek solutions f such that $|f| \leq 1$ for $R(z) > 0$.

$$|K e^{i\theta \log z / \log \lambda}| = |K| e^{-\theta (\arg z / \log \lambda)}.$$

¹² A. Denjoy, "Sur une classe des fonctions analytiques," *C. R. de l'acad. des sci. de Paris* (7 janvier 1929), pp. 140-142.

But since $R(z) > 0$, $|\arg z| < \pi/2$, and therefore $e^{-\theta(\arg z/\log \lambda)}$ is bounded for $R(z) > 0$. Thus if we choose $|K|$ sufficiently small, $|Ke^{i\theta \log z/\log \lambda}| \leq 1$ and therefore there is an infinity of functions satisfying 1) $f(\lambda z) = e^{i\theta} f(z)$ and 2) $|f| \leq 1$ for $R(z) > 0$ where θ is real and λ positive ($\neq 1$).

Similarly B) may be discussed. The functions

$$K + i\mu \log z/\log \lambda$$

where K is constant and $R(K)$ is sufficiently large satisfy all the requirements for f of B). Therefore returning to the equation (3.10), we find that, if U^* is elliptic or parabolic, there is always an infinity of solutions, and therefore, if w^* is not unique, w is not unique. Thus we have

THEOREM 3.2. *Let T of Theorem 3.1 be hyperbolic. Then the criterion of Denjoy is a necessary and sufficient condition for the uniqueness of w of Theorem 3.1.*

4. The principle of the harmonic majorant.¹³ Julia has stated and proved in his "Principes géométriques d'analyse" the following principle which he terms the "Principle of the Harmonic Majorant":

"Let $f(z)$ be a function of the complex variable z which satisfies the following conditions:

1) The function $f(z)$ is analytic and regular at every point of a region G_z . The modulus $|f(z)|$ is single-valued for $z \in G_z$.

2) There exists a function $u(z)$ harmonic and single-valued in G_z such that in the neighborhood of the boundary $\log |f(z)| - u(z)$ is less than every positive number; that is, for each point ξ of the boundary and for every positive ϵ , there exists a circle with center ξ such that at every point of G_z interior to this circle the inequality

$$\log |f(z)| - u(z) < \epsilon$$

is satisfied.

If these conditions are satisfied, $\log |f(z)| \leq u(z)$ at every point of G_z .

If the equality $\log |f(z)| = u(z)$ takes place at an interior point of G_z , $f(z)$ is of the form

$$e^{u(z)+iv(z)}$$

where $v(z)$ is a conjugate function of $u(z)$."

Let G_z be a doubly-connected region whose boundary consists of two disjoint continua. (We recall that a continuum is a closed set, not a single

¹³ G. Julia, *l. c.*

point, which is well-chained. The degenerate cases may be dismissed as trivial.) Further let us require that $f(z)$ be uniform for $z \subset G_z$ as well as that it satisfy the hypotheses of the Principle of the Harmonic Majorant. If e^{u+iv} is single-valued for $z \subset G_z$, then the inequality

$$(4.1) \quad \log |f(z)| \leq u(z)$$

is a "best possible" inequality and e^{u+iv} is an extremal function for the class of functions $f(z)$ which we are considering.

In general, e^{u+iv} , which we shall denote by $\phi(z)$, is not single-valued and the inequality (4.1) is strong. If we continue an element of $\phi(z)$ from a given point of G_z along a closed path in G_z containing in its interior one and only one of the continua which constitute the boundary of G_z back to the same given point, $\phi(z)$ changes to $e^{i\theta}\phi(z)$, which transformation we shall denote symbolically by

$$\phi(z) \rightarrow e^{i\theta}\phi(z) \quad (0 < \theta < 2\pi).$$

If $f(z)$ is analytic and single-valued for $z \subset G_z$ and satisfies (4.1), it follows that f/ϕ is analytic and has a single-valued modulus for $z \subset G_z$. Furthermore $|f/\phi| \leq 1$, and $f/\phi \rightarrow e^{-i\theta}f/\phi$. If $\psi(z)$ is analytic and has a single-valued modulus for $z \subset G_z$ and in addition 1) $|\psi| \leq 1$, 2) $\psi \rightarrow e^{-i\theta}\psi$, then $f = \psi\phi$ is analytic and single-valued for $z \subset G_z$ and $|f| \leq |\phi|$. Thus if we wish to calculate l. u. b. $|f(z_0)/\phi(z_0)|$ for $z_0 \subset G_z$ we may consider the equivalent problem of calculating l. u. b. $|\psi(z_0)|$ for $z_0 \subset G_z$ and for the class of functions $\{\psi(z)\}$ as defined above. It is clear from the definition of ψ that l. u. b. $|\psi(z_0)| < 1$ (for $\theta \not\equiv 0 \pmod{2\pi}$), which we shall denote by μ , is attained by some function $\psi^*(z)$ which belongs to the family $\{\psi(z)\}$. This is an immediate consequence of the theory of normal families.¹⁴ However we can go further. We shall show that μ is the limit of a monotonic non-increasing sequence which will be defined below and shall exhibit the totality of extremal functions which correspond to the bound μ .

To this end we shall introduce Poincaré's uniformisation function.¹⁵ Let G_z^∞ denote the universal covering surface of G_z and let $z = z(x)$, $|x| < 1$ denote the mapping function which maps the interior of the unit circle $|x| = 1$ one to one and conformally onto G_z^∞ such that $z(0) = 0$ and $z'(0) > 0$. It is well-known that $z(x)$ is automorphic under a cyclic group of trans-

¹⁴ P. Montel, *Leçons sur les familles normales* (Paris, 1927), p. 21.

¹⁵ H. Poincaré, *l. c.* For the properties of the mapping function $z(x)$ see G. Julia, *Leçons sur la représentation conforme des aires multiplement connexes* (Paris, 1934), Chap. 2 and 3.

formations $\{T^k\}$ where T is a hyperbolic transformation which corresponds to the passing from a given point ζ on a given sheet of G_z^∞ to that point of G_z^∞ which has the same geometric position as ζ and lies on the sheet which follows (or precedes) the sheet containing ζ .

Now consider the class of functions $w(x)$ defined for $|x| < 1$ by $w(x) = \psi(z(x))$. It is clear that $w(x)$ is analytic and single-valued for $|x| < 1$ and 1) $|w(x)| \leq 1$ for $|x| < 1$, 2) $w(T) = e^{-i\theta}w(x)$, and 3) $w(0) = \psi(z_0)$. Conversely, if $x(z)$ denotes the determination of the inverse function of $z(x)$, such that $x(z_0) = 0$, and if $w(x)$ is analytic for $|x| < 1$ and satisfies 1) $|w| \leq 1$ and 2) $w(T) = e^{-i\theta}w(x)$, then it follows that $\psi(z) = w(x(z))$ is defined and analytic for $z \in G_z$, has a single valued modulus there, and satisfies 1) $|\psi| \leq 1$, 2) $\psi \rightarrow e^{-i\theta}\psi$, and 3) $|\psi(z_0)| = |w(0)|$. Thus there is a complete one-to-one correspondence between the two classes of functions $\{\psi(z)\}$ and $\{w(x)\}$ and l.u.b. $|\psi(z_0)| = \text{l.u.b. } |w(0)|$ by virtue of 3) and the extremal functions of one class corresponds to the extremal functions of the other class under the conformal representation of G_z^∞ on $|x| < 1$. Therefore we shall consider in place of our original problem for ψ the equivalent problem of determining l.u.b. $|w(0)|$ where $w(x)$ is analytic for $|x| < 1$ and satisfies the two characteristic conditions 1) $|w| \leq 1$ and 2) $w(T) = e^{-i\theta}w(x)$.

Recalling that $\mu = \text{l.u.b. } |w(0)|$, we see that $0 < \mu < 1$, if $\theta \not\equiv 0 \pmod{2\pi}$. For one can construct simple examples of functions satisfying all the requirements of $w(x)$ which do not vanish for $x=0$. Without loss of generality we may assume that the extremal value μ is attained by a function $w(x)$ for which $w(0) > 0$, since the relation $w(T) = e^{-i\theta}w(x)$ is linear and homogeneous. Let μ_1 denote the largest positive number such that there exists a function $w_1(x) \subset \mathcal{E}$ for which $w_1(0) = \mu_1$, $w_1(T_0) = e^{-i\theta}\mu_1$, $\mu_1(T^{-1}0) = e^{i\theta}\mu_1$. For this value of μ_1 , 1) of Theorem 2.1 occurs when the notation is appropriately modified. For if 2) of Theorem 2.1 occurred, μ_1 could be replaced by a larger number since all the inequalities in 2) of Theorem 2.1 are strong. Thus μ_1 can be calculated directly from the relations 1) of Theorem 2.1. Similarly, let μ_2 denote the largest positive number such that there exists a function $w_2(x) \subset \mathcal{E}$ for which $w_2(T^k 0) = e^{-ki\theta}\mu_2$ ($k=0, \pm 1, \pm 2$). Here, too, 1) of Theorem 2.1 occurs, and therefore, μ_2 can be calculated algebraically from 1) of Theorem 2.1 and $\mu_1 \geq \mu_2$. In general, let μ_n denote the largest positive number such that there exists a function $w_n(x) \subset \mathcal{E}$ for which $w(T^k 0) = e^{-ki\theta}\mu_n$ ($k=0, \pm 1, \pm 2, \dots, \pm n$). Once again 1) of Theorem 2.1 occurs and μ_n can be calculated algebraically from the relations 1) of Theorem 2.1. The sequence $\{\mu_n\}$ is monotonic non-increasing and converges to a positive lower bound μ^* . Since $\mu \leq \mu_n$ for all n , it follows that $\mu \leq \mu^*$.

We shall prove that $\mu^* \leq \mu$ and therefore conclude that $\mu = \lim_{n \rightarrow \infty} \mu_n$. Let $w_n(x)$ denote the unique function of class \mathcal{E} for which

$$w_n(T^k 0) = e^{-ki\theta} \mu_n \quad (k = 0, \pm 1, \pm 2, \dots, \pm n).$$

It is clear that the sequence of functions $\{w_n\}$ forms a normal family and that any limit function $w^*(x)$ of the family belongs to class \mathcal{E} and satisfies the interpolation conditions $w^*(T^k 0) = e^{-ki\theta} \mu^*$ ($k = 0, \pm 1, \pm 2, \dots$). Let us recall that, as a consequence of Theorem 3.1, the existence of $w^* \in \mathcal{E}$ which satisfies the interpolation conditions $w^*(T^k 0) = e^{-ki\theta} \mu^*$, ($k = 0, \pm 1, \pm 2, \dots$) implies the existence of a function $w \in \mathcal{E}$ which not only satisfies the interpolation conditions of w^* but also the relation $w(T) = e^{-i\theta} w(x)$. Therefore $\mu^* \leq \mu$ and our assertion that $\mu = \lim_{n \rightarrow \infty} \mu_n$

is established. It is now easy to determine the totality of extremal functions for which $|w(0)| = \mu$. They are given by the formulas (3.8) and (3.10) where the notation is suitably modified. Thus returning to the class of functions $\{\psi\}$ which we have considered, we conclude

THEOREM 4.1. *For the function $\psi(z)$ defined above, we have l. u. b. $|\psi(z_0)| = \lim_{n \rightarrow \infty} \mu_n$. The totality of extremal functions $\psi^*(z)$ for which $|\psi^*(z_0)| = \lim_{n \rightarrow \infty} \mu_n = \mu$ is given by the totality of functions $w(x)$ for which $|w(0)| = \mu$. $|f(z_0)| \leq \mu |\phi(z_0)|$ where equality is attained for the functions $\psi^* \phi$ and only such functions.*

5. The principle of hyperbolic measure.¹⁶ We shall now enunciate the so-called "Principle of Hyperbolic Measure" and show how the methods which we have developed can be applied to study the extremal problems associated with this principle.

Let G_z and G_w be two regions which have each at least three boundary points and let $f(z)$ be an analytic function which can be continued throughout G_z such that its functional values lie in G_w . Let $t(w)$ map conformally the universal covering surface G_w^∞ of G_w on $|t| < 1$ and $x(z)$ map the universal covering surface G_z^∞ of G_z on $|x| < 1$. We form the function $t[f(z(x))] \equiv \phi(x)$ which can be extended throughout $|x| < 1$ and which takes on values interior to the circle $|t| = 1$. We denote the hyperbolic lengths¹⁷ of the four linear elements dx, dz, dw, dt by $d\sigma_x, d\sigma_z, d\sigma_w, d\sigma_t$ respectively so that in accordance with the invariance of these lengths under the transformations $x \rightarrow z$ and

¹⁶ Cf. note 4.

¹⁷ R. Nevanlinna, *Eindeutige Analytische Funktionen* (Berlin, 1936), Chap. 1.

$w \rightarrow t$, we have $d\sigma_x = d\sigma_z$ and $d\sigma_w = d\sigma_t$. We conclude from Pick's Theorem¹⁸ that $d\sigma_t \leq d\sigma_x$ and therefore that $d\sigma_w \leq d\sigma_z$. This is the *Principle of Hyperbolic Measure*.

If G_z is not simply-connected, and if we consider functions $f(z)$ which are single-valued for $z \in G_z$, then, in general, the inequality $d\sigma_w \leq d\sigma_z$ is to be replaced by the strong inequality $d\sigma_w < d\sigma_z$ and l. u. b. $d\sigma_{w_0}/d\sigma_{z_0} < 1$.

Let us suppose henceforth that G_z is a doubly-connected region the boundary of which consists of two disjoint continua. We shall consider functions which are analytic (save for possible poles) and single-valued for $z \in G_z$ such that $w = f(z) \in G_w$ where G_w is any region the boundary of which contains at least three points and $f(z_0) = w_0$ where z_0 is a given point of G_z and w_0 is a given point of G_w . Our problem is to determine effectively l. u. b. $d\sigma_{w_0}/d\sigma_{z_0}$.

As in the general statement of the Principle of Hyperbolic Measure, let $z(x)$ map $|x| < 1$ onto G_z^∞ one to one and conformally such that $z(0) = z_0$ and $z'(0) > 0$, and let $w(t)$ map $|t| < 1$ onto G_w if G_w is simply-connected, or if G_w is multiply-connected, onto G_w^∞ one to one and conformally such that $w(0) = w_0$ and $w'(0) > 0$. Then the function $\phi(x) \equiv t[f(z(x))]$, $|x| < 1$, where that determination of $t(w)$ is chosen for which $t(w_0) = 0$, as it has been defined above, has the properties $\phi(0) = 0$ and $|\phi| \leq |x|$ for $|x| < 1$ by Schwarz's Lemma. Furthermore we know that $z(x)$ is automorphic under a cyclic group of hyperbolic transformations $\{T^m\}$ generated from the hyperbolic transformation T . If G_w is simply-connected, the inverse function of $w(t)$ is single-valued; on the other hand, if G_w is multiply-connected, $w(t)$ is automorphic under a denumerable group of transformations $G_w[U_1, U_2, \dots]$ which are either hyperbolic or parabolic, and therefore $t(w)$, any determination of the inverse of $w(t)$, is a linear polymorphic function which has the law of transformation

$$t(w) \rightarrow U_k[t(w)]$$

when we continue $t(w)$ along a path in G_w^∞ from a given point on G_w^∞ to any point which has the same geometric position as the given point. With this fact in mind, let us study the possible functional relations which $\phi(x)$ may satisfy when x is replaced by Tx . It is evident that $z(T) = z(x)$; therefore $f(z(T))$ and $f(z(x))$ have the same geometric position on G_w^∞ and therefore

$$t[f(z(\tau))] = U_k[t(f(z(x)))]$$

¹⁸ G. Pick, "Ueber eine Eigenschaft der konforme Abbildung kreisformiger Bereiche," *Mathematische Annalen*, vol. 77 (1916), pp. 1-6.

where U_k is some substitution of the group G_w . Hence $\phi(x)$ satisfies a functional relation of the form

$$\phi(T) = U_k[\phi(x)]$$

where $U_k \subset G_w$. But not all substitutions $U_k \subset G_w$ are candidates. For if U_k is to be a candidate, we must have $\phi(x) = U_k^{-1}[\phi(Tx)]$ and we know from Schwarz's Lemma that $|\phi| \leq |x|$ for $|x| < 1$; therefore $|U_k^{-1}[\phi(T)]| \leq |x|$ for $|x| < 1$. Setting $x = T^{-1}0$ we find

$$|U_k^{-1}0| \leq |T^{-1}0| < 1.$$

Now there are only a finite number of substitutions U_k of the group G_w for which this is true.¹⁹ Furthermore it is conceivable that there need not exist a function ϕ analytic for $|x| < 1$ which vanishes at $x=0$ and satisfies the functional relation $\phi(T) = U_k[\phi(x)]$. (Let us remark that apart from the identical transformation the U_k are all hyperbolic or parabolic and therefore have fixed points on the unit circle.) By means of Theorem 3.1 we may eliminate those substitutions U_k which must be excluded, a fortiori, from our discussion. Let U_1, \dots, U_m denote those substitutions of G_w , finite in number, such that ϕ as it has been constructed, satisfies one (and only one) of the relations

$$\phi(T) = U_k[\phi(x)] \quad (k = 1, 2, \dots, m).$$

Conversely, let $\phi \in \mathcal{E}$ and furthermore satisfy 1) $\phi(0) = 0$, 2) $\phi(T) = U_k[\phi(x)]$ where U_k is one of the allowed substitutions. Then the function $f(z) \equiv w[\phi(x(z))]$ where the determination of $x(z)$ is so made that $x(z_0) = 0$, is defined and analytic throughout G_z . Furthermore $f(z)$ is single-valued. For, as $x(z)$ is continued along a path of G_z^∞ from a given point of G_z^∞ to a point which has the same geometric position as the given point, $x(z)$ is transformed into $Tx^n(z)$ and $\phi(x(z))$ is transformed into $U_k^n[\phi(x(z))]$. But $w(t)$ is automorphic under the group of substitutions G_w and therefore

$$w[U_k^n(\phi(x(z)))] = w[\phi(x(z))].$$

Therefore $f(z)$ is single-valued. It is evident that $w = f(z) \in G_w$ and $f(z_0) = w_0$. Thus the study of l. u. b. $d\sigma_{w_0}/d\sigma_{z_0}$ and the associated extremal functions is equivalent to the study of l. u. b. $\left. \frac{d\sigma_t}{d\sigma_x} \right|_{x=0}$ and the associated extremal functions where $t = \phi(x)$, $|x| < 1$ satisfies 1) $\phi(0) = 0$ and 2) $\phi(T) = U_k[\phi(x)]$ where U_k is one of the substitutions U_1, \dots, U_m

¹⁹ This is a consequence of the nature of the transformation U_k . See G. Julia, l. c. in note 15.

since $d\sigma_t = d\sigma_w$ and $d\sigma_z = d\sigma_x$. But let us note that since $\phi(0) = 0$, $d\sigma_x = |dx|$ and $d\sigma_t = |dt|$ and therefore

$$\text{l. u. b. } \left. \frac{d\sigma_t}{d\sigma_x} \right|_{x=0} = \text{l. u. b. } |\phi'(0)|.$$

From this it follows that $\text{l. u. b. } \left. \frac{d\sigma_t}{d\sigma_x} \right|_{x=0} = \max \mu^{(k)}$ where $\mu^{(k)} = \text{l. u. b. } |\phi'(0)|$ and ϕ satisfies the following conditions 1) $\phi \subset \mathcal{E}$, 2) $\phi(0) = 0$, 3) $\phi(T) = U_k[\phi(x)]$.

Let us determine $\mu^{(k)}$. We have shown in Section 3 that, if ϕ satisfies the conditions 1), 2), 3), it may be expressed in the form

$$(5.1) \quad \phi(x) = \frac{P(x) - Q(x)\phi_\infty(x)}{1 - S(x)\phi_\infty(x)}$$

where $\phi_\infty \subset \mathcal{E}$ and satisfies a functional relation of the form

$$(5.2) \quad \phi_\infty(T) = U_k^*[\phi_\infty(x)] \quad (U_k^* \text{ linear})$$

and where P, Q, S have the significance attributed to them in Section 3 with an appropriate modification of the notation. It is established in the Pick-Nevalinna theory²⁰ that $S(0) = 0$ and since $\phi(0) = 0$, we have $P(0) = Q(0) = 0$. Therefore $\phi'(0) = P'(0) - Q'(0)\phi_\infty(0)$ where $\phi_\infty \subset \mathcal{E}$ satisfies the relation (5.2). We have

$$(5.3) \quad \text{l. u. b. } |\phi'(0)| = \text{l. u. b. } |P'(0) - Q'(0)\phi_\infty(0)|.$$

We are now in a position to apply the methods which we have employed to discuss the extremal problem associated with the Principle of the Harmonic Majorant. Let $\mu_1^{(k)} = \max |P'(0) - Q'(0)\phi_\infty^{(1)}(0)|$ where $\phi_\infty^{(1)}(x) \subset \mathcal{E}$; $\phi_\infty^{(1)}$ can be determined directly. Let $\mu_2^{(k)} = \max |P'(0) - Q'(0)v_2^{(k)}|$ such that there exists a function $\phi_\infty^{(2)}(x) \subset \mathcal{E}$ for which

$$\phi_\infty^{(2)}(0) = v_2^{(k)}, \quad \phi_\infty^{(2)}(T_0) = U_k^* v_2^{(k)}, \quad \phi_\infty^{(2)}(T^{-1}0) = U_k^{*-1} v_2^{(k)}.$$

By 1) of Theorem 2.1 we are assured of the existence of $\phi_\infty^{(2)}(x)$ since there always exists a function $\phi_\infty \subset \mathcal{E}$ satisfying the relation (5.2). It is clear that $\mu_1^{(k)} \geq \mu_2^{(k)}$. In general, let $\mu_n^{(k)} = \max |P'(0) - Q'(0)v_n^{(k)}|$ such that there exists a function $\phi_\infty^{(n)} \subset \mathcal{E}$ for which

$$\phi_\infty^{(n)}(T^l 0) = U_k^{*l} v_n^{(k)} \quad (l = 0, \pm 1, \pm 2, \dots, \pm n).$$

It is clear that $\{\mu_n^{(k)}\}$ is a monotonic non-increasing sequence, such that $\mu_n^{(k)} \geq \mu^{(k)}$. As a consequence of exactly the same reasoning that we have

²⁰ J. L. Walsh, *l. c.*, p. 304.

employed in studying the extremal questions associated with the Principle of the Harmonic Majorant, we conclude that $\lim_{n \rightarrow \infty} \mu_n^{(k)} = \mu^{(k)}$.

Let $\mu = \max \mu^{(k)}$. It is now simple to determine the associated extremal functions. Let $\mu = |P'(0) - Q'(0)v|$ with the restriction that $|v| \leq 1$. Now, for those values of v so restricted that there exists a function $\phi_\infty \subset \mathcal{E}$ for which

$$\phi_\infty(T^l 0) = U^*_{\alpha^l v}(\mu^{(a)} = \max \mu^{(k)}), \quad (l = 0, \pm 1, \pm 2, \dots)$$

and only those values there correspond in accordance with Theorem 3.1 the associated extremal functions and Theorem 3.1 gives us the totality of such functions. By conformal transformation of the independent and dependent variables, we find the corresponding extremal functions in our original problem which is thus solved.

Let us remark that this result gives the exact value of the "Starrheitskonstant" Ω_0 of Aumann and Carathéodory²¹ as well as the associated extremal functions for doubly-connected regions.

6. The analogue of the Pick-Nevanlinna interpolation problem for doubly-connected regions. Theorems 3.1 and 3.2 virtually contain the solution to the following interpolation problem:

Let G_z be a doubly-connected region in the z -plane, the boundary of which consists of two disjoint continua. What is a necessary and sufficient condition that there exist a function $w(z)$ analytic for $z \in G_z$ which satisfies the following requirements: 1) $|w(z)| \leq 1$ for $z \in G_z$, 2) $w(z)$ is single-valued for $z \in G_z$, 3) $w(z_k) = w_k^{(0)}$ ($k = 1, 2, \dots, n$ or $k = 1, 2, \dots$) where $w_k^{(0)}$ are assigned complex numbers the moduli of which are not greater than unity and the z_k are distinct given points of G_z ?

If such a w exists, when is it unique?

If w is not unique, what is the totality of functions which satisfy the requirements 1), 2), 3)?

It is clear from our discussion that this problem is equivalent to the problem of Theorem 3.1 where U is the identical transformation and T is hyperbolic. Theorems 3.1 and 3.2 furnish the solution of this equivalent problem and therefore of the problem which we have just posed.

HARVARD UNIVERSITY.

²¹ G. Aumann and C. Carathéodory, *l. c.*

RAMANUJAN SUMS AND ALMOST PERIODIC FUNCTIONS.*

By M. KAC,¹ E. R. VAN KAMPEN and AUREL WINTNER.

Introduction. Several classical formal trigonometrical expansions of the analytic theory of numbers have recently been shown² to be periodic or almost periodic Fourier series of the functions which they represent. The object of the present paper is to prove a corresponding result for a class of multiplicative arithmetical sequences.

In particular, it will be shown that, for the functions to be considered, the celebrated formal trigonometric sums of Ramanujan³ are almost periodic Fourier expansions in the sense of Besicovitch. Hence, the Ramanujan coefficients will turn out to be Fourier averages which vanish for incommensurable values of the frequency parameter, the almost periodic function in question being always limit periodic (grenzperiodisch). It should be emphasized that the fact that the Ramanujan trigonometrical expansions turn out to be Fourier expansions leads without any further device to his explicit formulae, if one writes down the Fourier average representations of the coefficients.

Although the arithmetical functions $f(n)$ will be considered only for $n = 1, 2, \dots$, one can realize the usual assumption of the Besicovitch theory by placing $f(-n) = f(n)$ for $n = 1, 2, \dots$ and $f(0) = 0$ (the multiplicative character of f then remains preserved). It is understood that the class (B) of functions $f(n)$ which are defined for integers may be introduced either directly or by considering the step function $f(t)$ which has the value $f(n)$ for $n \leq t < n + 1$.

1. By a multiplicative function f is meant a sequence $f(n)$; $n = 1, 2, 3, \dots$ for which $f(n_1 n_2) = f(n_1) f(n_2)$ whenever $(n_1, n_2) = 1$ and $f(n) \neq 0$ for at least one n (so that $f(1) = 1$). Only those multiplicative $f(n)$ will be considered for which

$$(1) \quad f(n) = \prod_{p|n} f(p), \text{ i. e., } f(p) = f(p^2) = f(p^3) = \dots, (f(1) = 1),$$

* Received March 31, 1939.

¹ Fellow of the Parnas Foundation, Lwów, Poland.

² A. Wintner, *American Journal of Mathematics*, vol. 57 (1935), pp. 534-538; *Duke Mathematical Journal*, vol. 2 (1936), pp. 443-446; *American Journal of Mathematics*, vol. 59 (1937), pp. 629-634; P. Hartman and A. Wintner, *Travaux de l'Institut Math. de Tbilissi*, vol. 3 (1938), pp. 113-119; P. Hartman, *American Journal of Mathematics*, vol. 60 (1938), pp. 66-74; A. Wintner, *Revista de Ciencias* (Lima, 1939) (in press).

³ S. Ramanujan, *Collected Papers*, Cambridge University Press (1927), pp. 179-199.

where the p denote prime numbers. An $f(n)$ which satisfies (1) will be called strongly multiplicative. A classical instance of (1) is

$$(2) \quad f(n) = \frac{\phi(n)}{n} = \prod_{p|n} \frac{p-1}{p} = \prod_{p|n} \frac{\phi(p)}{p}; \quad (\phi = \text{Euler's function}).$$

For any $f(n)$ and for any positive integer k , put

$$(3) \quad f^{(k)}(n) = 1 \text{ or } f^{(k)}(n) = f(p_k) \text{ according as } n \not\equiv 0 \text{ or } n \equiv 0 \pmod{p_k},$$

where p_k is the k -th prime; and put

$$(4) \quad f_k(n) = \prod_{j=1}^k f^{(j)}(n); \text{ so that } f_k(n) = \prod_{p|n} f(p), \text{ where } p \leq p_k.$$

According to (3), the function $f^{(k)}(n)$ of n has the period p_k and possesses the Fourier expansion

$$(5) \quad f^{(k)}(n) = 1 + \frac{f(p_k) - 1}{p_k} \sum_{m=0}^{p_k-1} \exp(2\pi i \frac{m}{p_k} n),$$

which is, in fact, nothing but the formula of equidistant trigonometrical interpolation. According to (4), the function $f_k(n)$ of n has the period $P_k = p_1 p_2 \cdots p_{k-1} p_k$ and possesses, in view of (4) and (5), the Fourier expansion

$$(6) \quad f_k(n) = c_k + c_k \sum_{\substack{q|P_k \\ q > 1}} \sum_{\substack{(m, q)=1 \\ 1 \leq m < q}} \prod_{p|q} \frac{f(p) - 1}{f(p) - 1 + p} \cos(2\pi \frac{m}{q} n),$$

where $c_k = \prod_{p \leq p_k} \left(1 + \frac{f(p) - 1}{p}\right).$

2. For a function $g = g(n)$ defined for $n = 1, 2, 3, \dots$, put

$$(7) \quad M\{g\} = M\{g(n)\} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n g(m),$$

if this limit exists.

All considerations will be based on the following elementary lemma:

If a strongly multiplicative function $f(n)$ satisfies the condition

$$(8) \quad \sum \frac{|f(p) - 1|}{p} < \infty,$$

then the mean value $M\{f\}$ exists and

$$(9) \quad M\{f\} = \prod \left(1 + \frac{f(p) - 1}{p}\right).$$

In order to prove this, let $g(n)$ denote the multiplicative function which is defined as

$$g(n) = 0 \text{ or } g(n) = \prod_{p|n} \frac{f(p) - 1}{p}$$

according as n is not or is *quadratifrei*. Then, for every positive integer m ,

$$f(m) = \sum_{d|m} dg(d);$$

hence, $\sum_{m=1}^n f(m) = \sum_{m=1}^n \left[\frac{n}{m} \right] mg(m)$, and so

$$\sum_{m=1}^n f(m) = n \sum_{m=1}^n g(m) + O\left(\sum_{m=1}^n m |g(m)|\right).$$

Since the definition of $g(n)$ and the assumption (8) obviously imply the (absolute) convergence of the series $\sum_{m=1}^{\infty} g(m)$ to the sum represented by the product on the right of (9), it follows that in order to prove (9), it is sufficient to show that

$$O\left(\sum_{m=1}^n m |g(m)|\right) = o(n).$$

But the last relation is clear from the absolute convergence of the series $\sum_{m=1}^{\infty} g(m)$; so that the proof is complete.

The proof which we had originally for the above lemma was function-theoretical in nature. The above elementary approach was then suggested to us by Dr. Paul Erdős.

3. A corollary of (8)-(9) is that for a strongly multiplicative $f(n)$ one has

$$(10) \quad \lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{m=1}^n \frac{1}{mf(m)} = \prod \left(1 - \frac{1-f(p)}{pf(p)} \right), \text{ if } \sum \left| \frac{1-f(p)}{pf(p)} \right| < \infty.$$

In fact, on writing $\frac{1}{f(p)}$ for $f(p)$ in (8)-(9), one obtains (10), since

$$\text{if } \frac{1}{n} \sum_{m=1}^n a_m \rightarrow \alpha, \text{ then also } \frac{1}{\log n} \sum_{m=1}^n \frac{a_m}{m} \rightarrow \alpha.$$

Similarly, if $f(n)^\lambda$ denotes the λ -th power of $f(n)$ when either $f(n) > 0$ or λ is an integer, then

$$(11) \quad \lim_{n \rightarrow \infty} \frac{1}{n^{1+\lambda}} \sum_{m=1}^{\infty} m^\lambda f(m)^\lambda = \frac{1}{\lambda+1} \prod \left(1 - \frac{1-f(p)^\lambda}{p} \right),$$

if $\sum \frac{|1-f(p)^\lambda|}{p} < \infty$ and $\lambda > -1$

(and (10) may be thought of as the limiting case $\lambda = -1$). In order to prove (11), it is sufficient to replace $f(n)$ in (8)-(9) by the strongly multiplicative function $f(n)^\lambda$ and then apply the Abelian lemma:

$$\text{if } \sum_{m=1}^n a_m \sim \alpha n, \text{ then } (1 + \lambda) \sum_{m=1}^n m^\lambda a_m \sim \alpha n^{1+\lambda} \text{ for every } \lambda > -1.$$

As an illustration, consider the example (2); so that $f(p) = 1 - p^{-1}$. In this case, (10) is applicable and goes over into Landau's relation

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{m=1}^n \frac{1}{\phi(m)} = \frac{\zeta(2)\zeta(3)}{\zeta(6)}, \text{ since } \Pi \left(1 + \frac{1}{p(p-1)} \right) = \frac{\Pi(1-p^{-6})}{\Pi(1-p^{-3})\Pi(1-p^{-2})};$$

while (9) is applicable to any power of $\phi(n)/n$ and gives Schur's relation

$$M \left\{ \left(\frac{\phi(n)}{n} \right)^l \right\} = \Pi \left(1 - \frac{1}{p} + \frac{1}{p} \left(1 - \frac{1}{p} \right)^l \right)$$

for every real l (and, as seen from the proof of (9), for every complex l also).

4. For every strongly multiplicative, positive $f(n)$, let $f^+(n)$, $f^-(n)$ denote the strongly multiplicative, positive functions which at an arbitrary $n = p$ attain the values

$$f^+(p) = \text{Max}(1, f(p)) \text{ and } f^-(p) = \text{Min}(1, f(p)),$$

respectively. Then (2) shows that

$$(12_1) \quad f(n) = f^+(n)f^-(n); \quad (12_2) \quad 0 < f^-(n) \leq 1 \leq f^+(n);$$

while (4) clearly implies that

$$(13_1) \quad f^+(n) \geq f_k^+(n), \quad f^-(n) \leq f_k^-(n);$$

$$(13_2) \quad f - f_k = (f - f_k^-)f^+ + (f^+ - f_k^-)f_k^-.$$

Notice that either of the functions f_k^* is uniquely determined by f and k , i. e., that $(f^*)_k = (f_k)^*$.

Using these notations, it will be easy to deduce from (9) the following theorem:

Every strongly multiplicative, positive function $f(n)$ which satisfies (8) is almost periodic (B); furthermore,

$$(14) \quad M\{|f - f_k|\} \rightarrow 0, \text{ as } k \rightarrow \infty.$$

In fact, it is clear from (7) and (6) that $M\{f_k\} = c_k$. Since c_k in (6) was defined as the k -th partial product of the infinite product (9), it follows that

$$(14 \text{ bis}) \quad c_k = M\{f_k\} \rightarrow M\{f\}, \text{ as } k \rightarrow \infty.$$

Hence, (14) is certainly true if either $f(n) \geq f_k(n)$ or $f(n) \leq f_k(n)$ for every n and k . It follows therefore from (13₁) that

$$(15) \quad M\{|f^+ - f_k^+|\} \rightarrow 0 \text{ and } M\{|f^- - f_k^-|\} \rightarrow 0, \text{ as } k \rightarrow \infty.$$

But the function (6) of n is periodic for every f , hence also for f^+ ; so that either of the functions f_k^+ of n is periodic for every k . It follows therefore from (15) that either of the functions $f^+(n)$ is almost periodic (B). Since (12₂) shows that $f^-(n)$ is a bounded function, it follows from (12₁) that $f(n)$ is almost periodic (B).

In order to prove (14), notice first that, by (13₁) and (13₂),

$$(15 \text{ bis}) \quad M\{|f - f_k|\} \leq M\{(f_k^- - f^-)f^+\} + M\{(f^+ - f_k^+)f_k^-\}.$$

The sum $M + M$ on the right of (15 bis) may readily be written in the form $2M\{f_k^- f^+\} - M\{f\} - M\{f_k\}$. It follows therefore from (14 bis) and (15 bis) that in order to prove (14), it is sufficient to show that $M\{f_k^- f^+\} \rightarrow M\{f\}$ as $k \rightarrow \infty$. But this is obvious from (9) and from the definitions of f_k^- and f^+ .

5. The almost periodicity (B) of $f(n)$, proved in § 4, implies that the n -average $M\{f(n) \exp 2\pi i \lambda n\}$ exists for every real λ . It turns out that this Fourier coefficient vanishes for every irrational λ ; so that $f(n)$ is *limit periodic* (grenzperiodisch); more explicitly, the Fourier series (B) of $f(n)$ is

$$(16) \quad f(n) \sim M\{f\} + M\{f\} \sum_{q>1} \sum_m \sum_{p|q} \frac{f(p) - 1}{f(p) - 1 + p} \cos(2\pi \frac{m}{q} n),$$

where the first (exterior) summation is over all *quadratifrei* $q > 1$, while, if q is fixed, the index p runs through all prime divisors p of q , and m through the $\phi(q)$ values which satisfy $(m, q) = 1$ and $1 \leq m < q$.

In fact, (16) follows from (14), (14 bis) and (6), since P_k in (6) was defined as the product of the first k primes.

The restriction of the first summation index of (16) to *quadralfrei* $q > 1$ may be eliminated in the usual manner, if one introduces the Möbius function $\mu(r)$, where $r = 1, 2, 3, \dots$. In fact, (16) may then clearly be written in the form

$$(17) \quad f(n) \sim M\{f\} \sum_{r=1}^{\infty} \mu(r) c_r(n) \prod_{p|r} \frac{1-f(p)}{f(p)-1+p},$$

if $c_r(n)$ is an abbreviation for the finite sum

$$(18) \quad c_r(n) = \sum_m \cos\left(2\pi \frac{m}{r} n\right), \text{ where } (m, r) = 1 \text{ and } 1 \leq m < r.$$

Since the $\phi(r)$ angles which occur in the sum (18) are symmetrically placed, the sum which one obtains by writing \sin for \cos is 0; so that

$$(18 \text{ bis}) \quad c_r(n) = \sum_m \exp\left(2\pi i \frac{m}{r} n\right), \text{ where } (m, r) = 1 \text{ and } 1 \leq m < r.$$

Thus, the $c_r(n)$ are precisely the Ramanujan sums,⁴ and so the Fourier series (B) of $f(n)$ is identical with Ramanujan's formal trigonometric series for $f(n)$. The coefficients of the series

$$(19) \quad f(n) \sim \sum_{r=1}^{\infty} a_r c_r(n)$$

are

$$(20) \quad a_r = a_r(f) = M\{f\} \mu(r) \prod_{p|r} \frac{1-p}{f(p)-1+p} \quad (r = 1, 2, 3, \dots),$$

by (17); while the expansion functions (18) of (19) may be expressed⁵ in terms of the Euler ϕ -function and the Möbius μ -function as follows:

$$(21) \quad c_r(n) \phi\left(\frac{r}{t}\right) = \phi(r) \mu\left(\frac{r}{t}\right), \text{ where } t = (m, r).$$

6. According to (16), the frequencies (Fourier exponents) of the almost periodic function $f(n)$ are rational numbers between 0 and 1 (or, rather, between -1 and 1). Let the terms of the Fourier series (16) be ordered in the Ramanujan fashion (17)-(18), and suppose that each of them actually

⁴ S. Ramanujan, *loc. cit.*², pp. 180-181.

⁵ O. Hölder, Lichtenstein Memorial Volume, *Prace Matematyczno-Fizyczne*, vol. 43 (1936), pp. 13-23.

occurs, i. e., that none of the coefficients (20) of (19) vanishes. Then the frequencies of $f(n)$ are uniformly distributed on the interval $[0, 1]$ (or rather, $[-1, 1]$). This may be proved as follows:

Since $|\mu(m)| \leq 1$, while $\phi(m) \rightarrow \infty$ as $m \rightarrow \infty$, Hölder's formula (21) implies an observation of Ramanujan, according to which $c_r(n) = O(1)$ when either r is fixed and $n \rightarrow \infty$, or n is fixed and $r \rightarrow \infty$. In particular,

$$(21 \text{ bis}) \quad \lim_{r \rightarrow \infty} \frac{c_r(n)}{\phi(r)} = 0 \text{ for every fixed } n \geq 1.$$

Now, (21 bis) is equivalent to the equidistribution of the frequencies of (19).

In fact, let $S^{(r)}$ denote, for any fixed $r \geq 1$, the finite sequence

$$(22) \quad S^{(r)}: \frac{m_1^{(r)}}{r}, \frac{m_2^{(r)}}{r}, \dots, \frac{m_{\phi(r)}^{(r)}}{r}$$

of those fractions m/r whose numerator m satisfies the conditions $(m, r) = 1$ and $1 \leq m < r$. And let $\rho_r(x)$, $0 \leq x \leq 1$, denote the distribution function of the $\phi(r)$ fractions contained in $S^{(r)}$. Then it is clear from (18 bis) that the ratio occurring on the left of (21 bis) is the n -th Fourier-Stieltjes coefficient of $\rho_r(x)$, i. e., that

$$(22 \text{ bis}) \quad \int_0^1 \exp 2\pi i n x d\rho_r(x) = \frac{c_r(n)}{\phi(r)}; \quad (n \geq 1).$$

Thus, it is clear from the criterion of Weyl for equidistribution (mod 1), that the content of (21 bis) may be expressed as follows: The ordered infinite sequence of fractions which is obtained by writing $r = 1, 2, \dots$ in (22) is uniformly distributed on the interval $[0, 1]$. This fact, which is equivalent to a result of Pólya,⁶ may be obtained without the Fourier analysis (22 bis) of the sequence (22) also, and contains the corresponding fact concerning the ordered infinite sequence Farey sections.⁷

7. The considerations of § 4 and § 5 may be modified in such a way as to lead to (B^2) instead of to (B) . To this end, one merely has to replace the condition (8) by the pair of conditions

$$(23) \quad \sum \frac{|f(p) - 1|}{p} < \infty \text{ and } \sum \frac{|f(p)^2 - 1|}{p} < \infty.$$

⁶ Cf. G. Pólya and G. Szegő, *Aufgaben und Lehrsätze aus der Analysis*, chap. II, no. 188.

⁷ Cf. *loc. cit.*⁶, chap. II, no. 189.

In fact, a strongly multiplicative (real) function which satisfies (23) is almost periodic (B^2) and has the Fourier expansion (16) or (19); furthermore,

$$(24) \quad M\{(f - f_k)^2\} \rightarrow 0 \text{ as } k \rightarrow \infty,$$

and the Parseval relation takes the form

$$(25) \quad M\{f^2\} = \sum_{r=1}^{\infty} \phi(r) a_r^2.$$

In fact, if (23) is satisfied, then (4) shows that (9) is applicable to any of the three functions $f(n)^2$, $f_k(n)^2$, $f(n)f_k(n)$. Thus, the three averages $M\{f^2\}$, $M\{f_k^2\}$, $M\{ff_k\}$ exist and have the respective values

$$\prod_p \left(1 - \frac{1 - f(p)^2}{p}\right), \quad \prod_{p \leq p_k} \left(1 - \frac{1 - f(p)^2}{p}\right), \\ \prod_{p \leq p_k} \left(1 - \frac{1 - f(p)^2}{p}\right) \cdot \prod_{p > p_k} \left(1 - \frac{1 - f(p)^2}{p}\right).$$

Hence, $M\{f(n)^2\} + M\{f_k(n)^2\} - 2M\{f(n)f_k(n)\} \rightarrow 0$ as $k \rightarrow \infty$. This proves (24). Since $f_k(n)$ is, by § 1, a periodic function of n , it follows from (24) that $f(n)$ is almost periodic (B^2). Finally, (25) is clear from (17), since (19) and (18) show that every amplitude (20) occurs in (17) exactly $\phi(r)$ times.

As an illustration, consider the example (2). Then $f(p) = 1 - p^{-1}$; so that (23) is satisfied, and (20) shows that the coefficients (19) are

$$(26) \quad a_r = M\{f\} \mu(r) \prod_{p|r} (p^2 - 1)^{-1}; \quad (f(n) = \phi(n)/n).$$

THE JOHNS HOPKINS UNIVERSITY.

AN ASYMPTOTIC FORMULA FOR EXPONENTIAL INTEGRALS.* †

By PHILIP HARTMAN.

It is known¹ that if $f(x)$ is a function possessing a continuous second derivative in the interval $0 \leq x \leq 1$, then

$$(1) \quad \int_0^1 f(x) \exp(itx^\alpha) dx = \exp(\pi i/2\alpha) \Gamma(1 + \alpha^{-1}) f(0) t^{-1/\alpha} + O(t^{-2/\alpha}),$$

as $t \rightarrow +\infty$ for all $\alpha \geq 2$. Professor Wintner pointed out to me the problem suggested by the relation (1), namely, to determine conditions for the asymptotic formula (1) which are less restrictive than the assumption that $f(x)$ have a continuous second derivative, and to replace, at the same time, (1) by

$$(2) \quad \int_0^1 g(x) \exp(-sx^\delta) dx \sim Cg(0)s^{-1/\delta}, \quad |s| \rightarrow \infty,$$

where $s = \sigma + it$ is a complex variable. The object of this paper is to provide the answer to these questions.

It is clear that a necessary condition for (2) is that σ , the real part of s , should be non-negative. The case $\sigma = 0$ requires more stringent conditions than the case $\sigma > 0$. For this reason, the main results are stated in two theorems.

THEOREM 1. *If*

(i) $g(x)$ is of bounded variation in $0 \leq x \leq 1$, and

(ii) $\delta > 1$,

then

$$(3) \quad \delta \int_0^1 g(x) \exp(-sx^\delta) dx = \Gamma(\delta^{-1}) g(+0) s^{-1/\delta} + o(|s|^{-1/\delta}),$$

uniformly as $|s| \rightarrow \infty$ in the half-plane $|\arg s| \leq \pi/2$.

* Received December 14, 1938.

† Presented to the Society, February 25, 1939.

¹ The case $\alpha = 2$ was treated by O. Perron, "Über das infinitäre Verhalten der Koeffizienten einer gewissen Potenzreihe," *Archiv der Mathematik und Physik*, Series III, vol. 22 (1914), pp. 329-340. The formula (1) was proved for $\alpha \geq 2$ by A. Wintner, "On the asymptotic formulae of Riemann and of Laplace," *Proceedings of the National Academy of Sciences*, vol. 20 (1934), pp. 57-62.

THEOREM 2. *If*

- (i) *the integral ² of $g(x)$ over $0 \leq x \leq 1$ exists,*
- (ii) *$g(+0) = \lim_{x \rightarrow +0} g(x)$ exists, and*
- (iii) *$\delta > 1$,*

then (3) holds uniformly as $|s| \rightarrow \infty$ in the angular region $|\arg s| \leq \pi/2 - \epsilon$, where $\epsilon > 0$ is arbitrary. If condition (i) holds, but (ii), (iii) are replaced by

- (ii') *$|g(x) - g(+0)| |\log x|^{1/\delta} \rightarrow 0, x \rightarrow 0$, and*
- (iii') *$\delta > 0$,*

then (3) holds uniformly as $|s| \rightarrow \infty$ in the angular region $|\arg s| \leq \pi/2 - \epsilon$.

It has been recognized ³ that in the Laplace case, i. e. s real, $\Gamma(1 + \delta^{-1}) \times g(+0)s^{-1/\delta}$ is only the first term of an extended asymptotic formula for the integral in (3) if $g(x)$ possesses a sufficient number of derivatives. Actually, the same is true if s is not real. However, in the Riemann case, i. e. s purely imaginary, the remainder term cannot be better than $O(t^{-1})$. Furthermore, the condition that $g(x)$ possess a number of derivatives may be replaced by a much weaker condition. In this direction, we have the following corollaries:

COROLLARY 1. *Let $n \geq 1$ be an arbitrary integer; $\beta_k, c_k, k = 1, \dots, n$, arbitrary constants such that*

- (i) *$0 \leq \beta_1 < \beta_2 < \dots < \beta_n$*
- (ii) *$f(x) = \sum_{k=1}^n c_k x^{\beta_k} + h(x)x^{\beta_n}$, where $h(x)$ is of bounded variation in $0 \leq x \leq 1, h(+0) = 0$, and*
- (iii) *$\alpha > 1 + \beta_n$,*

then

$$(4) \quad \alpha \int_0^1 f(x) \exp(-sx^\alpha) dx = \sum_{k=1}^n c_k \Gamma[(1 + \beta_k)\alpha^{-1}] s^{-(1+\beta_k)/\alpha} + o(|s|^{-(1+\beta_n)/\alpha}),$$

holds uniformly as $|s| \rightarrow \infty$ in the half-plane $|\arg s| \leq \pi/2$.

COROLLARY 2. *Let $n \geq 1$ be an arbitrary integer; $\beta_k, c_k, k = 1, \dots, n$, arbitrary constants such that*

² In this paper, an integral over a finite interval is to be considered as an ordinary Lebesgue integral. An integral over an infinite interval is to be interpreted as an improper Riemann integral.

³ Cf. O. Perron, "Über die näherungsweise Berechnung von Funktionen grosser Zahlen," *Münchener Sitzungsberichte*, (1917), pp. 191-220; A. Haar, "Über Asymptotische Entwicklungen von Funktionen," *Mathematische Annalen*, vol. 96 (1926), pp. 69-107; A. Wintner, "Untersuchungen über Funktionen grosser Zahlen" *Mathematische Zeitschrift*, vol. 28 (1928), pp. 416-429; A. Wintner, *loc. cit.* 1.

$$(i) \ 0 \leq \beta_1 < \beta_2 < \cdots < \beta_n$$

$$(ii) \ f(x) = \sum_{k=1}^n c_k x^{\beta_k} + h(x) x^{\beta_n}, \text{ where the integral of } h(x) \text{ over } 0 \leq x \leq 1 \text{ exists,}$$

$$(iii) \ h(x) |\log x|^{(1+\beta_n)/\alpha} \rightarrow 0, \ x \rightarrow 0, \text{ and}$$

$$(iv) \ \alpha > 0,$$

then (4) holds uniformly as $|s| \rightarrow \infty$ in the angular region $|\arg s| \leq \pi/2 - \epsilon$.

These corollaries reduce to the corresponding theorem if $n = 1$, $\beta_1 = 0$, $c_1 = f(+0)$. It will be clear from the proof that a corollary analogous to the first part of Theorem 2 is true, i. e. if one replaces condition (iii), (iv) of Corollary 2 by

$$(iii') \ h(x) \rightarrow 0, \ x \rightarrow 0, \text{ and}$$

$$(iv') \ \alpha > 1 + \beta_n.$$

It may be noted that Corollary 1 for $n = 2$, $\beta_1 = 0$, $\beta_2 = 1$ is a slight improvement over (1) without any assumption as to the differentiability of $f(x)$. Also, if one does assume that $f(x)$ possesses m (> 0) continuous derivatives, an application of Corollary 2 gives an asymptotic formula with $(m + 1)$ terms, while earlier results⁴ in the Laplace case give a formula with only m terms.

First, the corollaries will be proved. By changing the integration variable from x to $x^{1/(1+\beta)}$, $\beta \geq 0$,

$$(5) \quad \int_0^1 x^\beta \exp(-sx^\alpha) dx = (1 + \beta)^{-1} \int_0^1 \exp[-sx^{\alpha/(1+\beta)}] dx.$$

Thus, one must consider integrals of the type

$$\int \exp(-sx^\gamma) dx, \quad \gamma = \alpha/(1 + \beta).$$

Now,

$$(6) \quad \int_0^1 \exp(-sx^\gamma) dx = \int_0^\infty \exp(-sx^\gamma) dx - \int_1^\infty \exp(-sx^\gamma) dx,$$

where the two integrals on the right of (6) exist if either

$$(7) \quad |\arg s| \leq \pi/2 \text{ and } \gamma > 1$$

or

$$(8) \quad |\arg s| \leq \pi/2 - \epsilon \text{ and } \gamma > 0.$$

⁴ Cf. A. Wintner, *loc. cit.* 1.

In either case ⁵

$$(9) \quad \int_0^{\infty} \exp(-sx^{\gamma}) dx = \Gamma(1 + \gamma^{-1}) s^{-1/\gamma} = \gamma^{-1} \Gamma(\gamma^{-1}) s^{-1/\gamma}.$$

In case (7), one has

$$(10) \quad \left| \int_1^{\infty} \exp(-sx^{\gamma}) dx \right| \leq 4\gamma^{-1} |s|^{-1}.$$

The appraisal (10) is obtained by changing the integration variable from x to $x^{1/\gamma}$ and applying the second mean value theorem to the resulting integral (over a finite interval)

$$\begin{aligned} \gamma^{-1} \int_1^b x^{(1-\gamma)/\gamma} \exp(-sx) dx \\ = \gamma^{-1} \int_1^{\xi} \exp(-sx) dx + \gamma^{-1} b^{(1-\gamma)/\gamma} \int_{\xi}^b \exp(-sx) dx. \end{aligned}$$

(It is understood that the second mean value theorem is applied separately to the real and imaginary parts of the integral and that, in the above formula and in the sequel, the following notation is used

$$(11) \quad \int^{\xi} (\cdot \cdot \cdot) dx = \int^{\xi} R(\cdot \cdot \cdot) dx + i \int^{\xi} I(\cdot \cdot \cdot) dx,$$

whenever ξ is a limit of integration.) By integrating, it is seen that the absolute value of each of the integrals on the right is less than $2|s|^{-1}$. The inequality (10) now follows by letting $b \rightarrow +\infty$.

On the other hand,⁶ in case (8)

$$(12) \quad \left| \int_1^{\infty} \exp(-sx^{\gamma}) dx \right| \leq C_N |s|^{-N},$$

where $N > 0$ is arbitrary and C_N depends only on N and γ . To prove (12), note that

$$\left| \int_1^{\infty} \exp(-sx^{\gamma}) dx \right| \leq \int_1^{\infty} \exp(-\sigma x^{\gamma}) dx,$$

where $s = \sigma + it$. By changing the integration variable from x to $\sigma^{1/\gamma}x$, the last integral becomes

$$(13) \quad \sigma^{-1/\gamma} \int \exp(-x^{\gamma}) dx = \sigma^{-1/\gamma} \int \exp(-x^{\gamma/2}) \exp(-x^{\gamma/2} + x^{\gamma/2}) dx,$$

⁵ Put $s = r \exp(i\theta)$; then

$$\int \exp(-sx^{\gamma}) dx = r^{-1/\gamma} \int \exp[-x^{\gamma} \exp(i\theta)] dx.$$

It can be shown by a straightforward application of Cauchy's integral theorem that in both cases (7) and (8)

$$\int_0^{\infty} \exp[-x^{\gamma} \exp(i\theta)] dx = \exp(-i\theta/\gamma) \int_0^{\infty} \exp(-x^{\gamma}) dx,$$

while the last integral is $\Gamma(1 + \gamma^{-1})$.

⁶ This appraisal is given by A. Wintner, *loc. cit.* 1.

where the limits of integration are $\sigma^{1/\gamma}$ and $+\infty$. It follows from (13), that

$$\left| \int_1^\infty \exp(-sx^\gamma) dx \right| < \sigma^{-1/\gamma} \exp(-\sigma^{1/2}) \int_0^\infty \exp(-x^\gamma + x^{\gamma/2}) dx,$$

from which one obtains (11).

Now, (5), (6), (9), (10), (12) imply

$$(14) \quad \alpha \int_0^1 \sum_{k=1}^n c_k x^{\beta_k} \exp(-sx^\alpha) dx \\ = \sum_{k=1}^n c_k \Gamma[(1+\beta_k)\alpha^{-1}] s^{-(1+\beta_k)/\alpha} + o(|s|^{-(1+\beta_n)/\alpha}),$$

uniformly as $|s| \rightarrow \infty$ if either

$$|\arg s| \leq \pi/2 \text{ and } \alpha > 1 + \beta_n$$

or

$$|\arg s| \leq \pi/2 - \epsilon \text{ and } \alpha > 0.$$

Thus, to complete the proof of the corollaries, it must be shown that

$$\int_0^1 h(x) x^\beta \exp(-sx^\alpha) dx = o(|s|^{-(1+\beta)/\alpha}).$$

By changing the integration variable from x to $x^{1/(1+\beta)}$, this becomes

$$\int_0^1 h[x^{1/(1+\beta)}] \exp(-sx^{\alpha/(1+\beta)}) dx = o(|s|^{-(1+\beta)/\alpha}).$$

On placing

$$g(x) = h[x^{1/(1+\beta)}],$$

it is seen that the Theorems 1 and 2 must be proved in the case $g(+0) = 0$.

Define the non-increasing function $m(|s|) = m(s)$ as follows

$$(15) \quad m(s) = \text{l. u. b. } |g(x)| \text{ for } 0 < x \leq |s|^{-1},$$

so that

$$m(s) \rightarrow 0, \quad |s| \rightarrow \infty.$$

Let $\phi(s) = \phi(|s|)$ be a non-decreasing function of $|s|$ which approaches ∞ with $|s|$ so slowly that

$$(16) \quad m[|s|^{1/\delta} \phi(s)^{-1}] \phi(s) \rightarrow 0, \quad |s| \rightarrow \infty.$$

For example, one may let

$$\phi(s) = \min[|s|^{1/2\delta}, m(|s|^{1/2\delta})^{-1/2}];$$

under the last set of conditions of Theorem 2, it may be supposed that⁷ in addition to (16)

⁷ It may be supposed that $1 \geq \delta > 0$, otherwise the second part of Theorem 2 is a special case of the first part. In this case, let

$$\psi(s) = \psi(|s|) = m(s) \log |s| \rightarrow 0, \quad |s| \rightarrow \infty.$$

The function $\phi(s)$ may be defined to be

$$\min[\psi(|s|^{1/2\delta-\eta})^{-1/2} \log^{1/\delta} |s|, |s|^{1/2\delta} \log^{1/\delta} |s|],$$

for a small constant $\eta > 0$.

$$(17) \quad \phi(s)^{\delta}/\log |s| \rightarrow \infty, \quad |s| \rightarrow \infty.$$

Now,

$$(18) \quad \int_0^1 g(x) \exp(-sx^{\delta}) dx = \delta^{-1} \int_0^1 g(x^{1/\delta}) x^{(1-\delta)/\delta} \exp(-sx) dx.$$

Consider the last integral from 0 to b , $0 < b \leq 1$,

$$\left| \int_0^b g(x^{1/\delta}) x^{(1-\delta)/\delta} \exp(-sx) dx \right| \leq m(b^{1/\delta}) \int_0^b x^{(1-\delta)/\delta} dx \\ = m(b^{1/\delta}) \delta b^{1/\delta}.$$

Thus, if one places

$$(19) \quad b = |s|^{-1} \phi(s)^{\delta},$$

it is seen from (16), that

$$(20) \quad \int_0^b g(x^{1/\delta}) x^{(1-\delta)/\delta} \exp(-sx) dx = o(|s|^{-1/\delta}),$$

uniformly as $|s| \rightarrow \infty$ in the half-plane $|\arg s| \leq \pi/2$. In order to appraise the integral on the right of (18) from b to 1, apply the second mean value theorem (to the monotone function $x^{(1-\delta)/\delta}$), one obtains

$$(21) \quad b^{(1-\delta)/\delta} \int_b^{\xi} g(x^{1/\delta}) \exp(-sx) dx + \int_{\xi}^1 g(x^{1/\delta}) \exp(-sx) dx,$$

where, cf. (11), $\xi = \xi(s) = (\xi_1, \xi_2)$ satisfies

$$(22) \quad b < \xi_1 < 1, \quad b < \xi_2 < 1.$$

The treatment of the integrals in (21) is essentially different for Theorem 1 and Theorem 2. Under the conditions of the first theorem, $g(x^{1/\delta})$ is of bounded variation and is, therefore, the difference of two non-decreasing functions. Thus, it may be supposed without loss of generality that $g(x^{1/\delta})$ is a bounded monotone function, so that the second mean value theorem may be applied to each of the integrals in (21). It follows that

$$(23) \quad \left| \int_b^1 g(x^{1/\delta}) x^{(1-\delta)/\delta} \exp(-sx) dx \right| \leq M b^{(1-\delta)/\delta} 8 |s|^{-1} + 8M |s|^{-1},$$

where $M = \text{l. u. b. } |g(x)|$ for $0 \leq x \leq 1$; so that by (19), and the fact that $\phi(s) \rightarrow \infty$ and $(1-\delta) < 0$, the integral in (23) is $o(|s|^{-1/\delta})$ uniformly as $|s| \rightarrow \infty$ in the half-plane $|\arg s| \leq \pi/2$. This completes the proof of Theorem 1 and Corollary 1.

Put $s = r \exp(i\theta)$; in the angular region $|\arg s| = |\theta| \leq \pi/2 - \epsilon$, one has $\cos \theta \geq c > 0$, for some constant $c = c_{\epsilon}$. Then for any $0 < p < q \leq 1$

$$(24) \quad \left| \int_p^q g(x^{1/\delta}) \exp(-sx) dx \right| \leq \int_p^q |g(x^{1/\delta})| \exp(-crx) dx.$$

Denote by S the set of points x in $0 \leq x \leq 1$ such that $|g(x^{1/\delta})| > 1$; and by T_{pq} the set of points in the interval $p \leq x \leq q$ which are not in S . Thus, if x is in T_{pq} , then $|g(x^{1/\delta})| \leq 1$; if x is in S , then $x > \eta$ for some constant $\eta > 0$ since $g(x) \rightarrow 0, x \rightarrow 0$. Therefore, by the first mean value theorem,

$$(25) \quad \int_{T_{pq}} |g(x^{1/\delta})| \exp(-cx) dx \leq \int_p^q \exp(-cx) dx < (cr)^{-1} \exp(-crp);$$

also

$$(26) \quad \int_S |g(x^{1/\delta})| \exp(-cx) dx \leq J \exp(-c\eta),$$

where

$$J = \int_0^1 |g(x^{1/\delta})| dx.$$

Combining (24), (25), (26),

$$(27) \quad \left| \int_p^q g(x^{1/\delta}) \exp(-sx) dx \right| < J \exp(-c\eta) + (cr)^{-1} \exp(-crp).$$

Thus, (19), (21), (22), (27) imply

$$(28) \quad \left| \int_h^1 g(x^{1/\delta}) x^{(1-\delta)/\delta} \exp(-sx) dx \right| < 2J[1 + r^{-(1-\delta)/\delta} \phi(s)^{1-\delta}] \exp(-c\eta) \\ + 2r^{-1/\delta} c^{-1} \phi(s)^{1-\delta} \exp[-c\phi(s)^\delta] + 2(cr)^{-1} \exp[-c\phi(s)^\delta].$$

The first term on the right of (28) is clearly $o(r^{-1/\delta}) = o(|s|^{-1/\delta})$. Since

$$\phi(s)^{1-\delta} \exp[-c\phi(s)^\delta] \rightarrow 0, \quad |s| \rightarrow \infty,$$

the second term is $o(r^{-1/\delta}) = o(|s|^{-1/\delta})$; while the last term is $o(r^{-1}) = o(|s|^{-1/\delta})$ if $\delta > 1$. Thus, the first part of Theorem 2 follows from (18), (20) and (28).

Under the conditions of the last part of Theorem 2, the last term on the right of (28) is also $o(r^{-1/\delta})$ even for $1 \geq \delta > 0$. For

$$r^{-1} \exp[-c\phi(s)^\delta] = r^{-1/\delta} \exp\{(-\log r)[c\phi(s)^\delta/\log r - (1-\delta)/\delta]\},$$

and the factor of $r^{-1/\delta}$ is $o(1)$ as $r \rightarrow \infty$ in virtue of (17). This completes the proof of Theorem 2 and Corollary 2.

QUEENS COLLEGE,
FLUSHING, NEW YORK.

ALMOST PERIODICITY AND THE REPRESENTATION OF INTEGERS AS SUMS OF SQUARES.*

By M. KAC.**

Let $r_s(n)$ be the number of different representations of n as a sum of s squares. Then

$$(1) \quad 1 + \sum_{n=1}^{\infty} r_s(n) q^n = (1 + 2 \sum_{n=1}^{\infty} q^{n^2})^s.$$

Hardy has given an analysis of the arithmetical properties of $r_s(n)$ based on the theory of elliptic ϑ -functions. The object of this note is to point out the close connection between the investigations of Hardy¹ and the theory of almost periodic sequences. In particular the "singular series" of Hardy and Littlewood turns out to be a formal Fourier expansion of $n^{1-2s}r_s(n)$. The main result of Hardy that for $5 \leq s \leq 8$ the sum of the "singular series" is precisely $n^{1-2s}r_s(n)$ will be shown to be equivalent to the statement that $n^{1-2s}r_s(n)$, where $5 \leq s \leq 8$, is a uniformly almost periodic sequence. The case $s=2$ is of particular interest, since $r_2(n)$ is² not even an almost periodic function of class (B) , although the Fourier coefficients exist and tend to 0. The investigations on almost periodicity of functions occurring in the analytic number theory and given by formal trigonometrical series, as originated by Wintner,³ have led, thus far, to functions which are almost periodic of the class (B^2) , at least. This may emphasize the interest of the situation mentioned above.

* Received May 11, 1939.

** Fellow of the Parnas Foundation, Lwów, Poland.

¹ G. H. Hardy, "On the representation of a number as the sum of any number of squares, and in particular of five," *Transactions of the American Mathematical Society*, vol. 21 (1920), pp. 255-284. Cf. also S. Ramanujan, "On certain arithmetical functions," *Collected papers of Srinivasa Ramanujan*, Cambridge (1927), pp. 136-162.

² A. S. Besicovitch, *Almost Periodic Functions*, Cambridge, 1932. In particular pp. 91-109.

³ A. Wintner, "On the asymptotic distribution of the remainder term of the prime-number theorem," *American Journal of Mathematics*, vol. 57 (1935), pp. 534-548; "The asymptotic behavior of the function $1/\zeta(1+it)$," *Duke Mathematical Journal*, vol. 2 (1936), pp. 443-446. Cf. also M. Kac, E. R. van Kampen and A. Wintner, "Ramanujans sums and almost periodic functions," *American Journal of Mathematics*, this number, pp. 107-114.

1. If $f(n)$ is a function defined for $n = 0, 1, 2, \dots$ and λ a real number, the averages $M\{f(n)\exp(2\pi i\lambda n)\}$, where

$$M\{g(n)\} = \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n g(j),$$

are called the Fourier coefficients of $f(n)$ if these averages exist. If these averages exist and if they vanish except for an at most enumerable set of λ -values, say for $\lambda = \lambda_1, \lambda_2, \dots$, the series

$$\sum_k \alpha_k \exp(-2\pi i\lambda_k n), \text{ where } \alpha_k = M\{f(n)\exp(2\pi i\lambda_k n)\},$$

will be called the Fourier series of $f(n)$, also when $f(n)$ is not almost periodic (B).

It will be shown that the Fourier coefficients of $f(n) = n^{1-\frac{1}{2}s}r_s(n)$ exist and that the Fourier series of $n^{1-\frac{1}{2}s}r_s(n)$ is the "singular series" of Hardy and Littlewood, namely

$$(2) \quad \rho_s(n) = \frac{\pi^{\frac{1}{2}s}}{\Gamma(\frac{1}{2}s)} \sum_{k=1}^{\infty} k^{-s} \sum_{(h,n)=1} (S_{hk})^s \exp(-2\pi i h n/k),$$

$$\text{where } S_{hk} = \sum_{j=0}^{k-1} \exp(2\pi i h j^2/k).$$

Suppose first that λ is irrational. Then

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{j=0}^{n-1} \exp(2\pi i \lambda j^2) = 0,$$

and so

$$(1-q)^{\frac{1}{2}s} (1 + 2 \sum_1^{\infty} q^{j^2} \exp(2\pi i \lambda j^2)) \rightarrow 0, \text{ as } q \rightarrow 1-0.$$

Thus, (1) implies that

$$(1-q)^{\frac{1}{2}s} (1 + \sum_1^{\infty} r_s(j) q^j \exp(2\pi i \lambda j)) \rightarrow 0, \text{ as } q \rightarrow 1-0.$$

Making use of a well known Tauberian theorem,⁴ one obtains

⁴ Cf. for instance J. Karamata, "Neuer Beweis und Verallgemeinerung einiger Tauberian Sätze," *Mathematische Zeitschrift*, vol. 33 (1931), pp. 294-299. The theorem cannot be applied directly since the coefficients are not positive. One can make use of the fact that $n^{-\frac{1}{2}s} \sum_1^n r_s(j) \rightarrow \pi^{\frac{1}{2}s} / \Gamma(\frac{1}{2}s + 1)$ and then to apply the theorem to the series $\sum r_s(n) (1 + \cos 2\pi \lambda n) q^n$ and $\sum r_s(n) (1 + \sin 2\pi \lambda n) q^n$. See also J. Karamata, "Sur les moyenne arithmétique des coefficients d'une série de Taylor," *Mathematica (Cluj)*, vol. 1 (1929), pp. 99-106.

$$n^{-\frac{1}{2}s} \sum_{j=0}^{n-1} r_s(j) \exp(2\pi i \lambda j) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

It follows now immediately that, for irrational λ ,

$$M\{n^{1-\frac{1}{2}s} r_s(n) \exp(2\pi i \lambda n)\} = 0.$$

Next, suppose that $\lambda = h/k$. Observing, as Hardy does, that

$$1 + 2 \sum_{j=1}^{\infty} q^{j^2} \exp(2\pi i h j^2 / k) \sim \pi^{\frac{1}{2}} \frac{S_{hk}}{k} (\log 1/q)^{-\frac{1}{2}} \sim \pi^{\frac{1}{2}} \frac{S_{hk}}{k} (1-q)^{-\frac{1}{2}}, \text{ as } q \rightarrow 1-0,$$

one obtains

$$1 + \sum_{j=1}^{\infty} r_s(j) q^j \exp(2\pi i h j / k) \sim \pi^{\frac{1}{2}s} \left(\frac{S_{hk}}{k}\right)^s (1-q)^{-\frac{1}{2}s}, \text{ as } q \rightarrow 1-0.$$

Applying again the Tauberian theorem,⁴ one arrives at

$$n^{-\frac{1}{2}s} \sum_{j=0}^{n-1} r_s(j) \exp(2\pi i h j / k) \rightarrow \frac{\pi^{\frac{1}{2}s}}{\Gamma(\frac{1}{2}s + 1)} \left(\frac{S_{hk}}{k}\right)^s, \text{ as } n \rightarrow \infty.$$

This obviously implies that

$$M\{n^{1-\frac{1}{2}s} r_s(n) \exp(2\pi i h n / k)\} = \frac{s}{2} \frac{\pi^{\frac{1}{2}s}}{\Gamma(\frac{1}{2}s + 1)} \left(\frac{S_{hk}}{k}\right)^s = \frac{\pi^{\frac{1}{2}s}}{\Gamma(\frac{1}{2}s + 1)} \left(\frac{S_{hk}}{k}\right)^s.$$

and the proof of the italicized statement is complete.

2. From the classical results concerning the Gaussian sums S_{hk} one readily deduces $|S_{hk}| \leq 2k^{\frac{1}{2}}$ and it is clear that for $s \geq 5$ the "singular series" (2) is absolutely convergent. Since it was already proved that the "singular series" is the Fourier series of $n^{1-\frac{1}{2}s} r_s(n)$, it follows that

$$n^{1-\frac{1}{2}s} r_s(n) = \rho_s(n)$$

holds if and only if $n^{1-\frac{1}{2}s} r_s(n)$ is uniformly almost periodic. The function $n^{1-\frac{1}{2}s} r_s(n)$ is not uniformly almost periodic for $s > 8$ and it is almost periodic for $5 \leq s \leq 8$. This is only a restatement of Hardy's results; it seems to be very difficult to prove or disprove elementarily the uniform almost periodicity of $n^{1-\frac{1}{2}s} r_s(n)$.

3. In the case $s = 3$ or $s = 4$ the "singular series" is not any more absolutely convergent and $n^{1-\frac{1}{2}s} r_s(n)$ is not uniformly almost periodic. Nevertheless it still has some properties of almost periodicity. In fact, it is almost

periodic (B^2). For simplicity, only the case $s = 4$ will be discussed. Let $\omega_j(n)$ be 1 or 0 according as j is or is not a divisor of n , and let $\gamma(n)$ be defined by

$$2^{\gamma(n)} | n, \quad 2^{\gamma(n)+1} \nmid n.$$

Jacobi's well known theorem concerning the representation of $r_4(n)$ in terms of $\sigma(n)$ may then be written as follows:

$$n^{-1}r_4(n) = 8 \frac{1 + 2\omega_2(n)}{2^{\gamma(n)+1} - 1} \frac{\sigma(n)}{n} = 8 \frac{1 + 2\omega_2(n)}{2^{\gamma(n)+1} - 1} \sum_{j=1}^{\infty} \frac{\omega_j(n)}{j}.$$

It can be easily verified that $M\{(n^{-1}\sigma(n) - \sum_{j=1}^N j^{-1}\omega_j(n))^2\} \rightarrow 0$, as $N \rightarrow \infty$.

This proves the almost periodicity (B^2) of $n^{-1}\sigma(n)$, since the finite sums $\sum_{j=1}^N j^{-1}\omega_j(n)$ are periodic. Observing that the ratio of $1 + 2\omega_2(n)$ and $2^{\gamma(n)+1} - 1$ is also almost periodic (B^2), one sees that so is $n^{-1}r_4(n)$.

From Jacobi's theorem one deduces by an elementary computation that $M\{n^{-2}r_4^2(n)\} = 420\zeta(3)$. On the other hand the Parseval relation gives

$$M\{n^{-2}r_4^2(n)\} = \pi^4 \sum_{k=1}^{\infty} k^{-8} \sum_{(h,k)=1} |S_{hk}|^8 \quad \text{where} \quad \pi^4 \sum_{k=1}^{\infty} k^{-8} \sum_{(h,k)=1} |S_{hk}|^8 = 420\zeta(3).$$

4. The Parseval relation is evidently valid in the case $5 \leq s \leq 8$, and it seems to be quite probable that it holds also for $s > 8$. This would mean that if $s > 8$, then $n^{1-\frac{s}{2}}r_s(n)$ is, though not uniformly almost periodic, at least almost periodic (B^2). Furthermore it would follow immediately that the inequality $|n^{1-\frac{s}{2}}r_s(n) - \rho_s(n)| > \epsilon$ holds for "almost all" integers (e. g. except for a sequence of integers of density 0) whatever is $\epsilon > 0$.

5. The case $s = 2$ is the most exceptional, which is due to the fact that $r_2(n)$ is 0 for "almost all" integers. $r_2(n)$ is evidently of class ($B^{1-\epsilon}$), since $M\{r_2(n)^{1-\epsilon}\} = 0$.

As mentioned in the introduction, $r_2(n)$ is not even almost periodic (B). The following simple proof of this statement was communicated to me by Dr. E. R. van Kampen.

Let $n = 2^{\alpha} p_1^{\beta_1} p_2^{\beta_2} \cdots q_1^{\gamma_1} q_2^{\gamma_2} \cdots$, where the p 's are primes $\equiv 3 \pmod{4}$ and the q 's primes $\equiv 1 \pmod{4}$. It is well known that

$$r_2(n) = \frac{1 + (-1)^{\beta_1}}{2} \frac{1 + (-1)^{\beta_2}}{2} \cdots (\gamma_1 + 1)(\gamma_2 + 1) \cdots.$$

Denote the product depending only on β 's by $\beta(n)$. It is easily seen that $\beta(n)$ is 0 for "almost all" integers and that $\beta(n)r_2(n) \equiv r_2(n)$. Suppose

now that $r_2(n)$ is almost periodic (B). According to a known theorem⁵ one could find a sequence of finite trigonometrical sums (which are in fact certain means of the partial sums of the "singular series") $W_k(n)$ approaching $r_2(n)$ in the mean, i. e., $M\{|r_2(n) - W_k(n)|\} \rightarrow 0$ as $k \rightarrow \infty$. Obviously

$$M\{|r_2(n) - \beta(n)W_k(n)|\} = M\{\beta(n) | r_2(n) - W_k(n) |\}$$

tends to 0 as $k \rightarrow \infty$. This implies a contradiction, since

$$\begin{aligned} M\{r_2(n)\} &= \pi, & M\{\beta(n) | W_k(n) |\} &= 0, \\ |r_2(n) - \beta(n)W_k(n)| &\geq r_2(n) - \beta(n) | W_k(n) |. \end{aligned}$$

6. It may be mentioned that the remarks of § 1 allow us to compute the limits of the expressions

$$n^{-\frac{1}{2}s} \sum_{j=1}^n r_s(jk)$$

as $n \rightarrow \infty$ and k is a fixed integer. In fact, let $\omega_k(n)$ have the same meaning as in § 3. Then $\omega_k(n) = k^{-1} \sum_{h=0}^{k-1} \exp(2\pi i h n / k)$, and so

$$\lim_{n \rightarrow \infty} n^{-\frac{1}{2}s} \sum_{j=1}^n r_s(jk) = \lim_{n \rightarrow \infty} n^{-\frac{1}{2}s} \sum_{j=1}^{nk} r_s(j) \omega_k(j) = \frac{\pi^{\frac{1}{2}s}}{k^{\frac{1}{2}s+1} \Gamma(\frac{1}{2}s + 1)} \sum_{h=0}^{k-1} (S_{hk})^s.$$

THE JOHNS HOPKINS UNIVERSITY.

⁵ *Loc. cit.*, 2) p. 105 (Theorem II).

ON THE EXPANSIONS OF CERTAIN MODULAR FORMS OF POSITIVE DIMENSION.*

By HERBERT S. ZUCKERMAN.¹

1. A definition of a modular form of positive dimension has been given in a paper by Rademacher and the author.² In that paper we found the Fourier expansions of those forms $F(\tau)$ which belong to the full modular group and which have only polar singularities at the parabolic point $\tau = i\infty$ when measured in the uniformizing variable $x = e^{2\pi i\tau}$. In the present paper we shall also restrict ourselves to functions which belong to the full group and which have only polar singularities at $\tau = i\infty$, but in the definition of a modular form we shall omit the restriction that $F(\tau)$ be analytic in the upper half-plane and shall merely assume that $F(\tau)$ has, as singularities in the fundamental region,³ at most a finite number of poles and possibly a polar singularity at $i\infty$.

This problem of determining expansions of modular forms having poles in the upper half-plane was partially considered by Hardy and Ramanujan.⁴ However they considered only forms of positive integral dimension which have no singularities at the parabolic points. The generalization to forms of real positive dimension presents no difficulties. We have only to introduce the roots of unity $\epsilon(a, b, c, d)$ and $e^{2\pi ia}$ in the transformation formulas (1.11) and (1.12) below, and to carry them through the analysis. However in order to take care of forms having singularities at $i\infty$ we have to evaluate certain integrals which Hardy and Ramanujan were able to eliminate by means of simple estimates. This part of the work is contained in sections 3, 4, 5, and 6.

In this paper we shall consider most of the integrals in the τ -plane rather than in the x -plane, where $x = e^{2\pi i\tau}$. The original path ω_N of section 2 is taken in the τ -plane so that we may avoid the poles of $F(\tau)$ which are easier to treat there than in the x -plane. After this point much of the work could

* Received May 16, 1939.

¹ Harrison Research Fellow.

² "On the Fourier coefficients of certain modular forms of positive dimension," *Annals of Mathematics*, vol. 39 (1938), pp. 433-462, especially section 1.

³ It is convenient to choose a particular fundamental region. We take the region $|\tau| \geq 1$, $-1/2 \leq \Re(\tau) \leq 1/2$, and throughout this paper we shall refer to it briefly as the fundamental region.

⁴ "On the coefficients in the expansions of certain modular functions," *Proceedings of the Royal Society, A*, vol. 95 (1919), pp. 144-155.

be done in either plane but by keeping in the τ -plane we are able to maintain a closer contact with the properties of modular forms and to eliminate certain rather artificial parts of the proof. For example, instead of using the mediants of the Farey series to break up the path of integration we are able to choose a set of points which are more closely connected with the form whose expansion we desire.

We write the transformation equations for our modular form as

$$(1.11) \quad F\left(\frac{a\tau + b}{c\tau + d}\right) = \epsilon(a, b, c, d) (-i(c\tau + d))^{-r} F(\tau), \quad c > 0,$$

$$(1.12) \quad F(\tau + 1) = e^{2\pi i a} F(\tau), \quad 0 \leq \alpha < 1,$$

where $|\epsilon(a, b, c, d)| = 1$ and where the branch of $(-i(c\tau + d))^{-r}$ is chosen as in the original definition of a modular form.

From (1.12) we have

$$(1.2) \quad e^{-2\pi i a(\tau+1)} F(\tau + 1) = e^{-2\pi i a} F(\tau),$$

and hence $e^{-2\pi i a} F(\tau)$ has a Fourier expansion for each region in which it is analytic. Since $F(\tau)$ has only a finite number of poles in the fundamental region, we can find a number A such that the only singularity of $F(\tau)$ with $\Im(\tau) \geq A$ is at $\tau = i\infty$. To simplify the later notation we shall always take for A a value ≥ 1 . Then, noting that $F(\tau)$ was restricted to have at most a polar singularity at $\tau = i\infty$, we see that we have a Fourier expansion

$$(1.3) \quad e^{-2\pi i a} F(\tau) = \sum_{n=-\mu}^{\infty} a_n e^{2\pi i n \tau} = \sum_{n=-\mu}^{\infty} a_n x^n, \quad a_{-\mu} \neq 0, \quad x = e^{2\pi i \tau},$$

which is valid for $\Im(\tau) \geq A$, $|x| \leq e^{-2\pi A}$.

For all τ in the upper half-plane we write

$$f(x) = e^{-2\pi i a} F(\tau).$$

Then, within the unit circle, $f(x)$ has a pole of order μ at $x = 0$, poles at each point corresponding to the poles of $F(\tau)$, and no other singularities. In section 2 we shall determine a closed curve C_N which lies within the unit circle, encloses the origin, and does not pass through any poles of $f(x)$. Then if y is a point within C_N and not a pole of $f(x)$ we have

$$\frac{1}{2\pi i} \int_{C_N} \frac{f(x)}{x - y} dx = f(y) + R(N),$$

where $R(N)$ is the sum of the residues of the function $f(x)/(x - y)$ at the poles of $f(x)$ which are enclosed by C_N . We then have

$$(1.4) \quad f(y) = \frac{1}{2\pi i} \int_{C_N} \frac{f(x)}{x-y} dx - R(N),$$

from which to obtain our expansion. This expansion will consist of two parts. The part arising from $R(N)$ corresponds to the Hardy and Ramanujan results while the part arising from the integral is analogous to the series obtained for forms having polar singularities at $i\infty$ and no other singularities in the upper half-plane. Despite this similarity it is not true that these two separate parts are each modular forms of dimension r belonging to the full group.

2. For the case in which $F(\tau)$ has no poles in the upper half-plane, the curve C_N could be chosen as the circle $x = \exp \{-2\pi N^{-2}\}$. However this curve is not suitable in our case since we must find one that avoids the poles of $f(x)$. The curve used by Hardy and Ramanujan can be used but it is geometrically more complicated than the one that will be used.

We first determine a path ω_N in the τ -plane and then take its image in the x -plane as C_N . For a positive integer N we let h_s/k_s be the s -th fraction in the Farey series⁵ of order N ,

$$(2.1) \quad \frac{0}{1}, \frac{1}{N}, \dots, \frac{h_{s-1}}{k_{s-1}}, \frac{h_s}{k_s}, \frac{h_{s+1}}{k_{s+1}}, \dots, \frac{N-1}{N}, \frac{1}{1},$$

and we consider the transformations

$$(2.2) \quad T_s = \begin{pmatrix} h_s & h_{s-1} \\ k_s & k_{s-1} \end{pmatrix},$$

where we define h_0 to be -1 and k_0 to be N . This transformation belongs to the modular group because of well known properties of Farey fractions. Under T_s the point $i\infty$ goes into h_s/k_s , 0 into h_{s-1}/k_{s-1} , and the line $\Re(\tau) = 0$ from 0 to $i\infty$ goes into the semicircle, through the points h_{s-1}/k_{s-1} and h_s/k_s , lying above the real axis. The first quadrant of the τ -plane is mapped into the area bounded by the semicircle and the real axis. If any poles of $F(\tau)$ lie on the line $\Re(\tau) = 0$ we detour around them with small semicircles extending into the right half-plane and deform the semicircles through h_s/k_s and h_{s-1}/k_{s-1} accordingly. These deformations will all extend downward into the semicircular area but will not reach the real axis.

These deformed semicircles join at the points h_s/k_s to yield a continuous path from $\tau = -1/N$ to $\tau = 1$. Finally we detour around the points h_s/k_s

⁵ Certain simple properties of the Farey series will be used without mention in this paper. For these properties and their proofs see E. Landau, *Vorlesungen über Zahlen-theorie* (1927), pp. 98-100.

along the line $\Im(\tau) = B$ from each semicircle to its neighbor. The constant B , which may vary with N , is to be taken small enough so that the line will meet all the semicircles. It is also to be such that this line does not contain any poles of $F(\tau)$ and we shall later add a further restriction.



The path ω_N for the value $N = 4$.

For ω_N we now choose the part of this path between the points $-1/(N+i)$ and $(i+N-1)/(N+i)$ which are the images of the point $\tau = i$ under the first and last of the T_s . The path ω_N is entirely above the real axis and it does not extend further above it than the largest semicircle. The radii of the semicircles are

$$\frac{1}{2} \left(\frac{h_s}{k_s} - \frac{h_{s-1}}{k_{s-1}} \right) = \frac{1}{2k_s k_{s-1}} \leq \frac{1}{2k_s(N+1-k_s)} < \frac{1}{2N},$$

and hence we have, for τ on ω_N ,

$$(2.3) \quad 0 < \Im(\tau) < \frac{1}{2N}.$$

The end points $-1/(N+i)$ and $(i+N-1)/(N+i)$ of ω_N differ by unity and hence the image C_N of ω_N in the x -plane is a closed curve. Also, by (2.3), we find, for x on C_N ,

$$\begin{aligned} |x| &= e^{-\pi \Im(\tau)} < 1, \\ 1 - |x| &= 1 - e^{-2\pi \Im(\tau)} < 1 - e^{-\pi N^{-1}}, \end{aligned}$$

and, therefore, C_N lies within the unit circle and approaches it as $N \rightarrow \infty$. We can now write equation (1.4) as

$$\begin{aligned} (2.4) \quad f(y) &= \int_{\omega_N} f(e^{2\pi i \tau}) (e^{2\pi i \tau} - y)^{-1} e^{2\pi i \tau} d\tau - R(N) \\ &= \int_{\omega_N} e^{2\pi i (1-\alpha) \tau} F(\tau) (e^{2\pi i \tau} - y)^{-1} d\tau - R(N). \end{aligned}$$

3. In order to evaluate the integral of (2.4) we break the path ω_N into parts. We let Ω_s be that part of ω_N which joins $h_{s-1}/k_{s-1} + iB$ to $h_s/k_s + iB$ for all s except the first and last, and for these values of s we let Ω_s be the corresponding remaining parts of ω_N . If we write

$$(3.11) \quad I_s = \int_{\Omega_s} e^{2\pi i(1-a)\tau} F(\tau) (e^{2\pi i\tau} - y)^{-1} d\tau,$$

we then have

$$(3.12) \quad \int_{\omega_Y} e^{2\pi i(1-a)\tau} F(\tau) (e^{2\pi i\tau} - y)^{-1} d\tau = \sum_s I_s,$$

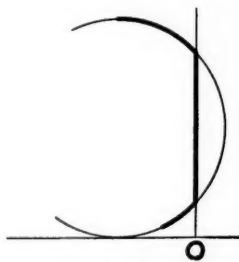
where \sum_s denotes the sum over all s for which there is a corresponding Farey fraction (2.1). In (3.11) we now make the change of variable

$$(3.21) \quad \tau = \frac{h_s\sigma + h_{s-1}}{k_s\sigma + k_{s-1}},$$

corresponding to the transformation (2.2). If τ lies on the part of Ω_s that does not consist of the line $\Im(\tau) = B$ then σ lies on the line $\Re(\sigma) = 0$ or on one of the detours around a pole of $F(\tau)$. If $\Im(\tau) = B$ then we have



The path Ω_s .



The path $\tilde{\Omega}_s$.

$$B = \Im\left(\frac{h_s\sigma + h_{s-1}}{k_s\sigma + k_{s-1}}\right) = \frac{\Im(\sigma)}{(k_s\Re(\sigma) + k_{s-1})^2 + k_s^2\Im(\sigma)^2},$$

$$\left(\Re(\sigma) + \frac{k_{s-1}}{k_s}\right)^2 + \left(\Im(\sigma) - \frac{1}{2k_s^2B}\right)^2 = \left(\frac{1}{2k_s^2B}\right)^2,$$

so σ lies on the circle with center at $-k_{s-1}/k_s + i/(2k_s^2B)$ and radius $1/(2k_s^2B)$. It can easily be verified that the points corresponding to the end points of Ω_s are

$$\sigma = -\frac{k_{s-1}}{k_s} + \frac{i}{k_s^2B}, \quad \sigma = \frac{-Bk_{s-1}^2}{Bk_s k_{s-1} + i},$$

except for the first and last values of s , in which case one end point corresponds to $\sigma = i$. Then as τ runs along Ω_s , σ runs along the circle from $-Bk_{s-1}^2/(Bk_s k_{s-1} + i)$ to the line $\Re(\sigma) = 0$; along this line, making the necessary detours, until it again meets the circle; and then along the circle to $-k_{s-1}/k_s + i/(k_s^2B)$. We shall call this path $\tilde{\Omega}_s$. By taking B sufficiently small we can keep the two points of intersection of the line and circle outside of the strip $1/A \leq \Im(\tau) \leq A$. We now have, using (1.11),

$$\begin{aligned}
 (3.22) \quad I_s &= \int_{\tilde{\Omega}_s} \exp \left\{ 2\pi i (1 - \alpha) \frac{h_s \sigma + h_{s-1}}{k_s \sigma + k_{s-1}} \right\} F \left(\frac{h_s \sigma + h_{s-1}}{k_s \sigma + k_{s-1}} \right) \\
 &\quad \times \left(\exp \left\{ 2\pi i \frac{h_s \sigma + h_{s-1}}{k_s \sigma + k_{s-1}} \right\} - y \right)^{-1} (k_s \sigma + k_{s-1})^{-2} d\sigma \\
 &= -\epsilon_s \int_{\tilde{\Omega}_s} \exp \left\{ 2\pi i (1 - \alpha) \frac{h_s \sigma + h_{s-1}}{k_s \sigma + k_{s-1}} \right\} \\
 &\quad \times \left(\exp \left\{ 2\pi i \frac{h_s \sigma + h_{s-1}}{k_s \sigma + k_{s-1}} \right\} - y \right)^{-1} (-i(k_s \sigma + k_{s-1}))^{-r-2} F(\sigma) d\sigma
 \end{aligned}$$

where we have used the abbreviation

$$(3.23) \quad \epsilon_s = \epsilon(h_s, h_{s-1}, k_s, k_{s-1}).$$

As we are later going to let N tend to infinity, we can suppose that it has been chosen so large that the curve C_N wholly contains the circle $|x| = (1 + |y|)/2$ for the particular point y , within the unit circle, that we are discussing. We then have the inequality

$$(3.3) \quad |x - y| \geq |x| - |y| \geq \frac{1 - |y|}{2}$$

for all x on the path C_N . The integrand of (3.22) can be written as

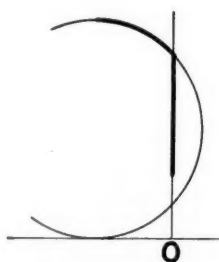
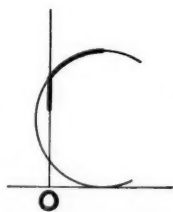
$$e^{2\pi i(1-\alpha)\tau} (x - y)^{-1} (-i(k_s \sigma + k_{s-1}))^{-r-2} F(\sigma),$$

where τ , which is given by (3.21), lies on Ω_s and $x = e^{2\pi i\tau}$ lies on C_N . Then, for the part of $\tilde{\Omega}_s$ in the strip $1/A \leq \Im(\sigma) \leq A$ we have $\Re(\sigma) \geq 0$ and hence

$$\begin{aligned}
 |e^{2\pi i(1-\alpha)\tau}| &= e^{-2\pi(1-\alpha)\Im(\tau)} < 1, \\
 |k_s \sigma + k_{s-1}|^2 &= (k_s \Re(\sigma) + k_{s-1})^2 + k_s^2 \Im(\sigma)^2 \geq k_{s-1}^2 + \frac{k_s^2}{A^2} \\
 &\geq \frac{1}{A^2} (k_{s-1}^2 + k_s^2) > \frac{1}{A^2} ((N - k_s)^2 + k_s^2) \geq \frac{N^2}{2A^2}.
 \end{aligned}$$

Also we see that the path is of finite length, is independent of N , and is free of poles, so $|F(\sigma)|$ has a bound on this path. Combining these results we see that the part of I_s due to this part of $\tilde{\Omega}_s$ is

$$O\left(\frac{N^{-r-2}}{1 - |y|}\right).$$

The path Ω_s' .The path Ω_s'' .

The remainder of $\tilde{\Omega}_s$ is in two parts. The upper part we call Ω_s' and the lower Ω_s'' . It is to be noted that Ω_1''' and Ω_s' for the largest admissible value of s are missing. For σ on Ω_s''' we make the change of variable $\rho = -1/\sigma$ and let Ω_s'' be the corresponding path in the ρ -plane. It can then be seen that Ω_s' consists of the path $\Re(\sigma) = 0$ from Ai to the circle $|\sigma + k_{s-1}/k_s - i/(2k_s^2 B)| = 1/(2k_s^2 B)$ and then the arc of this circle on to the point $-k_{s-1}/k_s + i/(k_s^2 B)$, while Ω_s'' consists of an arc of the circle $|\rho - k_s/k_{s-1} - i/(2k_{s-1}^2 B)| = 1/(2k_{s-1}^2 B)$ from $k_s/k_{s-1} + i/(k_{s-1}^2 B)$ to the line $\Re(\rho) = 0$ and then the segment of this line to Ai . These two paths lie entirely above the lines $\Im(\sigma) = \Im(\rho) = A$ and hence are free of the detours that we made to avoid the poles of $F(\tau)$. We now have

$$\begin{aligned}
 (3.4) \quad I_s = & -\epsilon_s \int_{\Omega_s'} \exp \left\{ 2\pi i (1 - \alpha) \frac{h_s \sigma + h_{s-1}}{k_s \sigma + k_{s-1}} \right\} \left(\exp \left\{ 2\pi i \frac{h_s \sigma + h_{s-1}}{k_s \sigma + k_{s-1}} \right\} - y \right)^{-1} \\
 & \times (-i(k_s \sigma + k_{s-1}))^{-r-2} F(\sigma) d\sigma \\
 & - \epsilon_s \int_{\Omega_s''} \exp \left\{ 2\pi i (1 - \alpha) \frac{h_{s-1} \rho - h_s}{k_{s-1} \rho - k_s} \right\} \left(\exp \left\{ 2\pi i \frac{h_{s-1} \rho - h_s}{k_{s-1} \rho - k_s} \right\} - y \right)^{-1} \\
 & \times \left(-i \left(-\frac{k_s}{\rho} + k_{s-1} \right) \right)^{-r-2} F\left(\frac{-1}{\rho}\right) \frac{d\rho}{\rho^2} + O\left(\frac{N^{-r-2}}{1-|y|}\right) \\
 = & H_s' + H_s'' + O\left(\frac{N^{-r-2}}{1-|y|}\right).
 \end{aligned}$$

Applying the transformation equation (1.11) to H_s'' we find

$$\begin{aligned}
 (3.51) \quad H_s'' = & -e^{\pi i r/2} \epsilon_0 \epsilon_s \int_{\Omega_s''} \exp \left\{ 2\pi i (1 - \alpha) \frac{h_{s-1} \rho - h_s}{k_{s-1} \rho - k_s} \right\} \\
 & \times \left(\exp \left\{ 2\pi i \frac{h_{s-1} \rho - h_s}{k_{s-1} \rho - k_s} \right\} - y \right)^{-1} (-i(k_{s-1} \rho - k_s))^{-r-2} F(\rho) d\rho,
 \end{aligned}$$

where

$$(3.52) \quad \epsilon_0 = \epsilon(0, -1, 1, 0).$$

On Ω_s'' we have $\Im(\rho) \geq A$ and hence the Fourier expansion (1.3) is valid.

Thus we may write

$$(3.53) \quad H_s'' = I_s'' + K_s'',$$

where

$$(3.54) \quad I_s'' = -e^{\pi i r/2} \epsilon_0 \epsilon_s \int_{\Omega_s''} \exp \left\{ 2\pi i (1-\alpha) \frac{h_{s-1}\rho - h_s}{k_{s-1}\rho - k_s} \right\} \\ \times \left(\exp \left\{ 2\pi i \frac{h_{s-1}\rho - h_s}{k_{s-1}\rho - k_s} \right\} - y \right)^{-1} \\ \times (-i(k_{s-1}\rho - k_s))^{-r-2} e^{2\pi i \alpha \rho} \sum_{\nu=1}^{\mu} a_{-\nu} e^{-2\pi i \nu \rho} d\rho,$$

$$(3.55) \quad K_s'' = -e^{\pi i r/2} \epsilon_0 \epsilon_s \int_{\Omega_s''} \exp \left\{ 2\pi i (1-\alpha) \frac{h_{s-1}\rho - h_s}{k_{s-1}\rho - k_s} \right\} \\ \times \left(\exp \left\{ 2\pi i \frac{h_{s-1}\rho - h_s}{k_{s-1}\rho - k_s} \right\} - y \right)^{-1} \\ \times (-i(k_{s-1}\rho - k_s))^{-r-2} e^{2\pi i \alpha \rho} \sum_{n=0}^{\infty} a_n e^{2\pi i n \rho} d\rho.$$

If ρ is on Ω_s'' then $\tau = (h_{s-1}\rho - h_s)/(k_{s-1}\rho - k_s)$ is on ω_N and $x = e^{2\pi i \tau}$ is on C_N and we may write (3.55) in the form

$$K_s'' = -e^{\pi i r/2} \epsilon_0 \epsilon_s \int_{\Omega_s''} \exp \{ 2\pi i (1-\alpha) \tau \} (x-y)^{-1} \\ \times (-i(k_{s-1}\rho - k_s))^{-r-2} e^{2\pi i \alpha \rho} \sum_{n=0}^{\infty} a_n e^{2\pi i n \rho} d\rho, \\ |K_s''| = \left| \int_{\omega_s''} \exp \{ 2\pi i (1-\alpha) \tau \} (x-y)^{-1} (-i(k_{s-1}\rho - k_s))^{-r-2} e^{2\pi i \alpha \rho} \sum_{n=0}^{\infty} a_n e^{2\pi i n \rho} d\rho \right|$$

where ω_s'' is the part of ω_N that τ runs over as ρ runs over Ω_s'' . From the geometry of the path Ω_s'' and properties of the Farey fractions we have

$$|k_{s-1}\rho - k_s|^2 \geq |k_{s-1}A i - k_s|^2 = k_{s-1}^2 A^2 + k_s^2 \\ \geq k_{s-1}^2 + k_s^2 > (N - k_s)^2 + k_s^2 \geq \frac{N^2}{2}.$$

Also we have $\Im(\rho) \geq A$, the inequality (2.3) for τ , and (3.3) for x , and hence

$$(3.56) \quad K_s'' = O\left(\frac{1}{1-|y|} N^{-r} e^{-2\pi \alpha A} \sum_{n=0}^{\infty} |a_n| e^{-2\pi n A} \int_{\omega_s''} |d\tau|\right) \\ = O\left(\frac{N^{-r}}{1-|y|} \int_{\omega_s''} |d\tau|\right).$$

In a similar way we find

$$(3.61) \quad H_s' = I_s' + K_s',$$

where

$$(3.62) \quad I_s' = -\epsilon_s \int_{\Omega_s'} \exp \left\{ 2\pi i (1-\alpha) \frac{h_s\sigma + h_{s-1}}{k_s\sigma + k_{s-1}} \right\} \left(\exp \left\{ 2\pi i \frac{h_s\sigma + h_{s-1}}{k_s\sigma + k_{s-1}} \right\} - y \right)^{-1} \\ \times (-i(k_s\sigma + k_{s-1}))^{-r-2} e^{2\pi i \alpha \sigma} \sum_{\nu=1}^{\mu} a_{-\nu} e^{-2\pi i \nu \sigma} d\sigma,$$

and

$$(3.63) \quad K'_s = O\left(\frac{N^{-r}}{1-|y|} \int_{\omega'_s} |d\tau|\right).$$

Now ω'_s and ω''_s are both parts of Ω_s and they do not overlap. Hence we may combine (3.56) and (3.63) to obtain

$$(3.64) \quad K'_s + K''_s = O\left(\frac{N^{-r}}{1-|y|} \int_{\Omega_s} |d\tau|\right).$$

4. We now consider the integrals I'_s and I''_s of (3.62) and (3.54). In I'_s we make the change of variable

$$z = -i(k_s\sigma + k_{s-1}), \quad \sigma = i\frac{z}{k_s} - \frac{k_{s-1}}{k_s},$$

and have

$$(4.11) \quad I'_s = -i\epsilon_s \frac{1}{k_s} \int \exp\left\{2\pi i(1-\alpha)\left(\frac{h_s}{k_s} + \frac{i}{k_s z}\right)\right\} \\ \times \left(\exp\left\{2\pi i\left(\frac{h_s}{k_s} + \frac{i}{k_s z}\right)\right\} - y\right)^{-1} z^{-r-2} \\ \times \exp\left\{2\pi i\alpha\left(-\frac{k_{s-1}}{k_s} + i\frac{z}{k_s}\right)\right\} \sum_{\nu=1}^{\mu} a_{-\nu} \exp\left\{-2\pi i\nu\left(-\frac{k_{s-1}}{k_s} + i\frac{z}{k_s}\right)\right\} dz,$$

where the path of integration consists of the line $\Im(z) = -k_{s-1}$ from the point $ik_s - ik_{s-1}$ to the point at which this line meets the circle corresponding to the circular part of Ω'_s , and then an arc of this circle on to the point $1/(k_s B)$.

In I''_s we set

$$z = -i(k_{s-1}\rho - k_s), \quad \rho = i\frac{z}{k_{s-1}} + \frac{k_s}{k_{s-1}},$$

and find

$$(4.12) \quad I''_s = -ie^{\pi i r/2} \epsilon_0 \epsilon_s \frac{1}{k_{s-1}} \int \exp\left\{2\pi i(1-\alpha)\left(\frac{h_{s-1}}{k_{s-1}} + \frac{i}{k_{s-1} z}\right)\right\} \\ \times \left(\exp\left\{2\pi i\left(\frac{h_{s-1}}{k_{s-1}} + \frac{i}{k_{s-1} z}\right)\right\} - y\right)^{-1} z^{-r-2} \\ \times \exp\left\{2\pi i\alpha\left(\frac{k_s}{k_{s-1}} + \frac{iz}{k_{s-1}}\right)\right\} \sum_{\nu=1}^u a_{-\nu} \exp\left\{-2\pi i\nu\left(\frac{k_s}{k_{s-1}} + \frac{iz}{k_{s-1}}\right)\right\} dz,$$

where the path of integration consists of a circular arc and a straight line joining the points $1/(k_{s-1}B)$ and $ik_s + k_{s-1}$.

We now wish to combine I'_s and I''_{s+1} into a single integral. Since their paths of integration abut at the point $1/(k_s B)$ it will be necessary only to transform the integrand of I''_{s+1} to show that it is the same as that of I'_s . It is because of this that we chose the path ω_N in such a way that the first and last Ω_s were incomplete. The effect of that choice is that I''_1 and I'_s

for the largest admissible value of s are missing and hence when we sum the I_s of (3.11) to obtain the integral of (2.4) we are to sum the expression $(I'_s + I''_{s+1})$ over the s corresponding to all Farey fractions (2.1) with the exception of the last one, $1/1$.

In order to transform the integrand of I''_{s+1} we first note that from the property

$$h_s k_{s-1} - h_{s-1} k_s = 1, \quad h_{s+1} k_s - h_s k_{s+1} = 1,$$

of the Farey fractions, we have

$$(4.21) \quad h_{s+1} k_{s-1} - h_{s-1} k_{s+1} = \frac{h_{s-1} + h_{s+1}}{h_s} = \frac{k_{s-1} + k_{s+1}}{k_s}.$$

Using this, the transformation equations (1.11) and (1.12), and the definitions (3.23) and (3.52), we find

$$(4.22) \quad F\left(\frac{h_s \tau + h_{s-1}}{k_s \tau + k_{s-1}}\right) = F\left(\frac{h_{s+1}(-(\tau + h_{s+1} k_{s-1} - h_{s-1} k_{s+1})^{-1}) + h_s}{k_{s+1}(-(\tau + h_{s+1} k_{s-1} - h_{s-1} k_{s+1})^{-1}) + k_s}\right) \\ = \epsilon_{s+1} \left(-i \frac{k_s \tau + k_{s-1}}{\tau + h_{s+1} k_{s-1} - h_{s-1} k_{s+1}}\right)^{-\tau} F\left(\frac{-1}{\tau + h_{s+1} k_{s-1} - h_{s-1} k_{s+1}}\right) \\ = \epsilon_{s+1} \epsilon_0 e^{\pi i \tau / 2} (-i(k_s \tau + k_{s-1}))^{-\tau} F(\tau + h_{s+1} k_{s-1} - h_{s-1} k_{s+1}) \\ = \epsilon_{s+1} \epsilon_0 e^{\pi i \tau / 2} \exp\{2\pi i \alpha(h_{s+1} k_{s-1} - h_{s-1} k_{s+1})\} (-i(k_s \tau + k_{s-1}))^{-\tau} F(\tau),$$

and

$$(4.23) \quad F\left(\frac{h_s \tau + h_{s-1}}{k_s \tau + k_{s-1}}\right) = \epsilon_s (-i(k_s \tau + k_{s-1}))^{-\tau} F(\tau),$$

and hence we have

$$(4.3) \quad e^{\pi i \tau / 2} \epsilon_0 \epsilon_{s+1} \exp\left\{2\pi i \alpha \frac{k_{s+1}}{k_s}\right\} \exp\left\{-2\pi i \nu \frac{k_{s+1}}{k_s}\right\} \\ = \epsilon_s \exp\left\{-2\pi i \alpha \frac{k_{s-1}}{k_s}\right\} \exp\left\{-2\pi i \nu \left(-\frac{k_{s-1}}{k_s}\right)\right\}.$$

Equation (4.3) is obtained by comparing (4.22) with (4.23) and using (4.21). Using this result we may write (4.12) as

$$I''_{s+1} = -i \epsilon_s \frac{1}{k_s} \int \exp\left\{2\pi i(1-\alpha)\left(\frac{h_s}{k_s} + \frac{i}{k_s z}\right)\right\} \\ \times \left(\exp\left\{2\pi i\left(\frac{h_s}{k_s} + \frac{i}{k_s z}\right)\right\} - y\right)^{-1} z^{-r-2} \\ \times \exp\left\{2\pi i \alpha \left(-\frac{k_{s-1}}{k_s} + \frac{iz}{k_s}\right)\right\} \sum_{\nu=1}^u a_{-\nu} \exp\left\{-2\pi i \nu \left(-\frac{k_{s-1}}{k_s} + \frac{iz}{k_s}\right)\right\} dz,$$

which we now combine with (4.11) to get

$$(4.4) \quad I'_s + I''_{s+1} = -i \epsilon_s \frac{1}{k_s} \int \exp\left\{2\pi i(1-\alpha)\left(\frac{h_s}{k_s} + \frac{i}{k_s z}\right)\right\} \\ \times \left(\exp\left\{2\pi i\left(\frac{h_s}{k_s} + \frac{i}{k_s z}\right)\right\} - y\right)^{-1} z^{-r-2} \\ \times \exp\left\{2\pi i \alpha \left(-\frac{k_{s-1}}{k_s} + \frac{iz}{k_s}\right)\right\} \sum_{\nu=1}^u a_{-\nu} \exp\left\{-2\pi i \nu \left(-\frac{k_{s-1}}{k_s} + \frac{iz}{k_s}\right)\right\} dz.$$

The path of integration of (4.4) extends from $Ak_s - ik_{s-1}$ to $Ak_s + ik_{s+1}$. The exact form of this path need not be known. It is sufficient to note that if we trace back through the changes of variable that we have made, we find that as z runs over this path then

$$x = \exp \left\{ 2\pi i \left(\frac{h_s}{k_s} + \frac{i}{k_s z} \right) \right\}$$

runs over a part of C_N . By (3.3) we then have

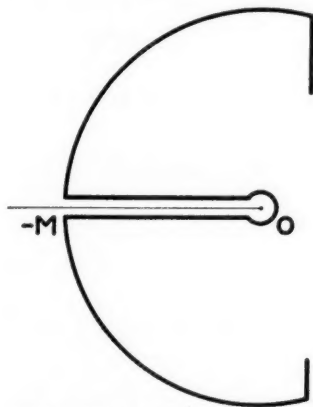
$$\left| \exp \left\{ 2\pi i \left(\frac{h_s}{k_s} + \frac{i}{k_s z} \right) \right\} \right| > |y|,$$

and we can expand the denominator of the integrand of (4.4) in a uniformly convergent geometric series of non-negative powers of y . Doing this and interchanging the orders of summation and integration we obtain

$$(4.5) \quad I'_s + I''_{s+1} = -i\epsilon_s \frac{1}{k_s} \sum_{\nu=1}^{\mu} \sum_{n=0}^{\infty} a_{-\nu} \exp \left\{ 2\pi i \left(-(n+\alpha) \frac{h_s}{k_s} + (\nu-\alpha) \frac{k_{s-1}}{k_s} \right) \right\} y^n \\ \times \int z^{-\nu-2} \exp \left\{ \frac{2\pi}{k_s} (\nu-\alpha) z + (n+\alpha) \frac{1}{z} \right\} dz.$$

5. The integrand in (4.5) has no singularities except at the points $z=0$ and $z=\infty$ so we can deform the path of integration. If $n+\alpha > 0$ we cut the z -plane along the negative real axis and take the following path:

$\Re(z) = Ak_s$ from $Ak_s - ik_{s-1}$ to the point at which $|z| = M$,
 $|z| = M$ from the point at which $\Re(z) = Ak_s$ to the point $-M$,
 a loop from $-M$ around the origin and back to $-M$ (on the upper border),
 $|z| = M$ from $-M$ to the point at which $\Re(z) = Ak_s$,
 $\Re(z) = Ak_s$ from the point at which $|z| = M$ to ik_{s+1} .



The path of integration
for $I'_s + I''_{s+1}$.

Calling the integrals along these paths J_1, J_2, J_3, J_4, J_5 , respectively, we have

$$\begin{aligned} |J_2 + J_4| &\leq \int_{|z|=M} M^{-r-2} \exp \left\{ \frac{2\pi}{k_s} \left((v-\alpha)Ak_s + (n+\alpha)\frac{1}{M} \right) \right\} |dz| \\ &\leq 2\pi \exp \left\{ \frac{2\pi}{k_s} \left((v-\alpha)Ak_s + (n+\alpha)\frac{1}{M} \right) \right\} M^{-r-1}, \end{aligned}$$

and hence $J_2 + J_4 \rightarrow 0$ as $M \rightarrow \infty$. Also, on the path of J_1 we have

$$\begin{aligned} z &= Ak_s - y_i, \quad k_{s-1} \leq y \leq M, \\ \Re(z) &= Ak_s, \quad \Re\left(\frac{1}{z}\right) = \frac{Ak_s}{A^2k_s^2 + y^2} \leq \frac{Ak_s}{k_s^2 + (N-k_s)^2} \leq \frac{2Ak_s}{N^2}, \\ |z|^2 &\geq A^2k_s^2 + k_{s-1}^2 \geq k_s^2 + (N-k_s)^2 \geq \frac{N^2}{2}, \end{aligned}$$

and hence

$$\begin{aligned} |J_1| &\leq \int_{Ak_s - ik_{s-1}}^{Ak_s - i\infty} 2^{r/2} N^{-r} \frac{1}{|z|^2} \exp \left\{ \frac{2\pi}{k_s} \left((v-\alpha)Ak_s + (n+\alpha)\frac{2Ak_s}{N^2} \right) \right\} |dz| \\ &= O\left(N^{-r} \exp\{4A\pi N^{-2}(n+\alpha)\} \int_{Ak_s - ik_{s-1}}^{Ak_s - i\infty} |z|^{-2} |dz|\right) \\ &= O\left(N^{-r} \exp\{4A\pi N^{-2}(n+\alpha)\} \int_{(Ak_s - ik_{s-1})^{-1}}^0 |dw|\right), \end{aligned}$$

where, in the last integral, we have set $w = 1/z$ and where the path of integration is along an arc of the circle tangent to the imaginary axis at the origin and passing through the point $(Ak_s - ik_{s-1})^{-1}$. The length of this path is at most $\pi/2$ times the length of the chord:

$$\frac{\pi}{2} |Ak_s - ik_{s-1}|^{-1} = \frac{\pi}{2} (A^2k_s^2 + k_{s-1}^2)^{-1/2} \leq \frac{\pi}{\sqrt{2}} N^{-1},$$

and therefore we have

$$J_1 = O(N^{-r-1} \exp\{4A\pi N^{-2}(n+\alpha)\}),$$

and, similarly,

$$J_5 = O(N^{-r-1} \exp\{4A\pi N^{-2}(n+\alpha)\}).$$

Finally we have ⁶

$$\begin{aligned} \lim_{M \rightarrow \infty} J_3 &= \int_{-\infty}^{(0+)} z^{-r-2} \exp \left\{ \frac{2\pi}{k_s} \left((v-\alpha)z + (n+\alpha)\frac{1}{z} \right) \right\} dz \\ &= \frac{(2\pi)^{r+1}}{k_s^{r+1}} (v-\alpha)^{r+1} \int_{-\infty}^{(0+)} t^{-r-2} \exp \left\{ t + \frac{4\pi^2(n+\alpha)(v-\alpha)}{k_s^2 t} \right\} dt \\ &= 2\pi i \left(\frac{v-\alpha}{n+\alpha} \right)^{(r+1)/2} I_{r+1} \left(\frac{4\pi}{k_s} \sqrt{(n+\alpha)(v-\alpha)} \right). \end{aligned}$$

⁶ G. N. Watson, *Theory of Bessel Functions* (1922), p. 181, (1).

Combining these results and letting M tend to infinity we then find that the integral of (4.5) has the value

$$(5.1) \quad 2\pi i \left(\frac{\nu - \alpha}{n + \alpha} \right)^{(r+1)/2} I_{r+1} \left(\frac{4\pi}{k_s} \sqrt{(n + \alpha)(\nu - \alpha)} \right) + O(N^{-r-1} \exp\{4A\pi N^{-2}(n + \alpha)\}).$$

If $n + \alpha = 0$ we have $n = \alpha = 0$. It has been shown in the first paper⁷ referred to in section 1 that $\alpha = 0$ implies $r = \text{an integer}$. The integral can then be evaluated in this case, as was done in that paper, without cutting the z -plane. The result is that (5.1) may still be used provided we use the convention of that paper regarding its meaning in this case.

We can now combine (2.4), (3.11), (3.4), (3.53), (3.61), (3.64), (4.5), and (5.1) to obtain

$$(5.2) \quad f(y) = 2\pi \sum_s \left\{ \sum_{\nu=1}^{\mu} a_{-\nu} \frac{1}{k_s} \sum_{n=0}^{\infty} \epsilon_s \exp \left\{ 2\pi i \left(-(n + \alpha) \frac{h_s}{k_s} + (\nu - \alpha) \frac{k_{s-1}}{k_s} \right) \right\} \right. \\ \times \left(\left(\frac{\nu - \alpha}{n + \alpha} \right)^{(r+1)/2} I_{r+1} \left(\frac{4\pi}{k_s} \sqrt{(n + \alpha)(\nu - \alpha)} \right) \right. \\ \left. \left. + O(N^{-r-1} \exp\{4A\pi N^{-2}(n + \alpha)\}) \right\} y^n \right. \\ \left. + O \left(\frac{N^{-r}}{1 - |y|} \int_{\Omega_s} |d\tau| \right) + O \left(\frac{N^{-r-2}}{1 - |y|} \right) \right\} - R(N).$$

The error terms in (5.2) reduce to

$$(5.3) \quad O \left(\sum_s \left\{ \frac{1}{k_s} \sum_{n=0}^{\infty} N^{-r-1} \exp\{4A\pi N^{-2}(n + \alpha)\} |y|^n \right. \right. \\ \left. \left. + \frac{N^{-r}}{1 - |y|} \int_{\Omega_s} |d\tau| + \frac{N^{-r-2}}{1 - |y|} \right\} \right) \\ = O(N^{-r} \exp\{4A\pi N^{-2}\alpha\} \sum_s N^{-1} k_s^{-1} \sum_{n=0}^{\infty} (\exp\{4A\pi N^{-2}\} |y|)^n) \\ + O \left(\frac{N^{-r}}{1 - |y|} \sum_s \left(\int_{\Omega_s} |d\tau| + N^{-2} \right) \right) \\ = O(N^{-r} \exp\{4A\pi N^{-2}\alpha\} \sum_s (1 - \exp\{4A\pi N^{-2}\} |y|)^{-1} N^{-1} k_s^{-1}) \\ + O \left(\frac{N^{-r}}{1 - |y|} \left(\int_{\omega_N} |d\tau| + \sum_s N^{-2} \right) \right).$$

Now it is easily seen that the length of the path ω_N has an upper bound independent of N . Also we have

$$\sum_s N^{-2} \leq \sum_s N^{-1} k_s^{-1} \leq \sum_{k=1}^N \sum_{h=0}^{k-1} N^{-1} k^{-1} = \sum_{k=1}^N N^{-1} = 1,$$

⁷ *Loc. cit.*, footnote 2, sections 6 and 7.

and therefore the error terms (5.3) have the estimate

$$O(N^{-r} \exp\{4A\pi N^{-2}\alpha\} (1 - \exp\{4A\pi N^{-2}\} |y|)^{-1}),$$

and (5.2) may now be written as

$$(5.4) \quad f(y) = 2\pi \sum_s \sum_{\nu=1}^{\mu} a_{-\nu} \frac{1}{k_s} \sum_{n=0}^{\infty} \epsilon_s \exp \left\{ 2\pi i \left(-(n+\alpha) \frac{h_s}{k_s} + (\nu-\alpha) \frac{k_{s-1}}{k_s} \right) \right\} \\ \times \left(\frac{\nu-\alpha}{n+\alpha} \right)^{(r+1)/2} I_{r+1} \left(\frac{4\pi}{k_s} \sqrt{(n+\alpha)(\nu-\alpha)} \right) y^n \\ - R(N) + O(N^{-r} \exp\{4A\pi N^{-2}\alpha\} (1 - \exp\{4A\pi N^{-2}\} |y|)^{-1}).$$

6. If we let $N \rightarrow \infty$ in (5.4) the error term will tend to zero and we shall have the desired expansion of $f(y)$. In order to perform this we must free (5.4) of its dependence on the Farey series of order N . That is, we must replace the term

$$\epsilon_s \exp \left\{ 2\pi i (\nu - \alpha) \frac{k_{s-1}}{k_s} \right\}$$

by an equivalent expression involving h_s and k_s but not h_{s-1} and k_{s-1} . It is convenient at this point to omit the subscript s , writing h and k for h_s and k_s but still using h_{s-1} and k_{s-1} . We first define h' to be any solution of the congruence

$$(6.1) \quad hh' \equiv -1 \pmod{k},$$

from which we have, at once,

$$h' + k_{s-1} \equiv 0 \pmod{k}.$$

By (3.23) and (1.11) we have

$$F\left(\frac{h\tau + h_{s-1}}{k\tau + k_{s-1}}\right) = \epsilon_s (-i(k\tau + k_{s-1}))^{-r} F(\tau).$$

On the other hand we use (6.1), (1.11), (1.12), and the fact that h_{s-1}/k_{s-1} and h/k are successive Farey fractions to get

$$F\left(\frac{h\tau + h_{s-1}}{k\tau + k_{s-1}}\right) = F\left(\frac{h(\tau + (k_{s-1} + h')/k) - (1 + hh')/k}{k(\tau + (k_{s-1} + h')/k) - h'}\right) \\ = \epsilon \left(h, -\frac{1 + hh'}{k}, k, -h' \right) (-i(k\tau + k_{s-1}))^{-r} \\ \times \exp \left\{ 2\pi i \alpha \frac{k_{s-1} + h'}{k} \right\} F(\tau).$$

Comparing these two results we have

$$\epsilon_s = \epsilon \left(h, -\frac{1 + hh'}{k}, k, -h' \right) \exp \left\{ 2\pi i \alpha \frac{k_{s-1} + h'}{k} \right\},$$

and hence, since ν is an integer and k divides $(h' + k_{s-1})$,

$$\epsilon_s \exp \left\{ 2\pi i(\nu - \alpha) \frac{k_{s-1}}{k} \right\} = \epsilon \left(h, -\frac{1 + hh'}{k}, k, -h' \right) \exp \left\{ -2\pi i(\nu - \alpha) \frac{h'}{k} \right\}.$$

Using this we may now write (5.4) in the form

$$(6.21) \quad f(y) = 2\pi \sum_{k=1}^N \sum_{\nu=1}^{\mu} a_{-\nu} \frac{1}{k} \sum_{n=0}^{\infty} A_{k,\nu}(n) \left(\frac{\nu - \alpha}{n + \alpha} \right)^{(r+1)/2} \\ \times I_{r+1} \left(\frac{4\pi}{k} \sqrt{(\nu - \alpha)(n + \alpha)} \right) y^n - R(N) \\ + O(N^{-r} \exp\{4A\pi N^{-2}\} (1 - \exp\{4A\pi N^{-2}\} |y|)^{-1}),$$

where

$$(6.22) \quad A_{k,\nu}(n) = \sum_{\substack{0 \leq h \leq k \\ (h,k)=1}} \epsilon \left(h, -\frac{1 + hh'}{k}, k, -h' \right) \\ \times \exp \left\{ -2\pi i \left((n + \alpha) \frac{h}{k} + (\nu - \alpha) \frac{h'}{k} \right) \right\}.$$

If $N \rightarrow \infty$ the first term of the right member of (6.21) becomes an infinite series which is easily seen to be convergent. The second term becomes $R(\infty)$ which also stands for an infinite series. Since the error term becomes zero, this second series then converges, provided its terms are summed in the proper order. By $R(\infty)$ we shall mean the infinite series whose value is $\lim_{N \rightarrow \infty} R(N)$. Then we have:

THEOREM 1. *If $F(\tau)$ is a modular form of positive dimension r , having, as singularities in the fundamental region, at most a finite number of poles and a polar singularity at $i\infty$, then we have the expansion*

$$(6.3) \quad F(\tau) = e^{2\pi i a \tau} f(e^{2\pi i \tau}), \\ f(y) = 2\pi \sum_{k=1}^{\infty} \sum_{\nu=1}^{\mu} a_{-\nu} \frac{1}{k} \sum_{n=0}^{\infty} A_{k,\nu}(n) \left(\frac{\nu - \alpha}{n + \alpha} \right)^{(r+1)/2} \\ \times I_{r+1} \left(\frac{4\pi}{k} \sqrt{(\nu - \alpha)(n + \alpha)} \right) y^n - R(\infty),$$

where $A_{k,\nu}(n)$ is defined by (6.22) and the remaining constants are determined by the transformation equations (1.11) and (1.12) and by the "principal part" of the Fourier expansion (1.3).

7. We shall now obtain the series representation for $R(\infty)$ in the case in which $F(\tau)$ has a single simple pole in the interior of the fundamental region and no other singularities in this region except a possible polar singularity at $i\infty$. The value of $R(\infty)$ in other cases can be obtained in a similar manner, with obvious modifications.

We let $F(\tau)$ have its simple pole at the point $\tau = \sigma$ and suppose that its residue there is R . Then, expanding $F(\tau)$ about this point, we have

$$(7.11) \quad F(\tau) = \frac{R}{\tau - \sigma} + \sum_{n=0}^{\infty} c_n (\tau - \sigma)^n.$$

Now if τ lies in the triangle which is obtained from the fundamental region by the transformation

$$(7.12) \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then $(-d\tau + b)/(c\tau - a)$ lies in the fundamental region and we have, by (1.11),

$$(7.13) \quad F(\tau) = \epsilon(-d, b, c, -a)^{-1} (-i(c\tau - a))^r F\left(\frac{-d\tau + b}{c\tau - a}\right).$$

In this triangle we have a single determination of the branch of $(-i(c\tau - a))^r$ and therefore $F(\tau)$ is analytic except for the parabolic points and the points $(a\sigma + b)/(c\sigma + d)$. For τ near the latter points we combine (7.11) and (7.13) to obtain the expansion

$$\begin{aligned} F(\tau) &= \epsilon(-d, b, c, -a)^{-1} (-i(c\tau - a))^r \\ &\quad \times \left\{ \frac{R}{(-d\tau + b)/(c\tau - a) - \sigma} + \sum_{n=0}^{\infty} c_n \left(\frac{-d\tau + b}{c\tau - a} - \sigma \right)^n \right\} \\ &= \epsilon(-d, b, c, -a)^{-1} (-i(c\tau - a))^r \left\{ \frac{-(c\tau - a)R}{c\sigma + d} \right. \\ &\quad \left. + \frac{1}{\tau - (a\sigma + b)/(c\sigma + d)} + \sum_{n=0}^{\infty} c_n \left(\frac{-d\tau + b}{c\tau - a} - \sigma \right)^n \right\}, \end{aligned}$$

from which we see that $F(\tau)$ has a simple pole at $(a\sigma + b)/(c\sigma + d)$ with the residue

$$\begin{aligned} &-\epsilon(-d, b, c, -a)^{-1} \left(-i \left(c \frac{a\sigma + b}{c\sigma + d} - a \right) \right)^r \left(c \frac{a\sigma + b}{c\sigma + d} - a \right) \frac{R}{c\sigma + d} \\ &= -\epsilon(-d, b, c, -a)^{-1} (-i(c\sigma + d))^{-r-2} R. \end{aligned}$$

Corresponding to this pole, $f(x) = e^{-2\pi i a \tau} F(\tau)$ will have a simple pole at the point

$$(7.21) \quad x = \exp \left(2\pi i \frac{a\sigma + b}{c\sigma + d} \right),$$

with the residue

$$(7.22) \quad -2\pi i \exp \left\{ 2\pi i (1 - \alpha) \frac{a\sigma + b}{c\sigma + d} \right\} \epsilon(-d, b, c, -a)^{-1} (-i(c\sigma + d))^{-r-2} R.$$

Not all transformations (7.12) will give distinct points (7.21). Since σ is not a vertex of the fundamental region we shall get each point

$(a\sigma + b)/(c\sigma + d)$ once and only once if we take the identity transformation and all transformations (7.12) with $c \geq 1$. Two of these points, $(a\sigma + b)/(c\sigma + d)$ and $(a_1\sigma + b_1)/(c_1\sigma + d_1)$, will yield the same point (7.21) in the x -plane if and only if their difference is an integer t ,

$$\frac{a_1\sigma + b_1}{c_1\sigma + d_1} = \frac{a\sigma + b}{c\sigma + d} + t = \frac{(a + tc)\sigma + (b + td)}{c\sigma + d}.$$

Therefore we can take each pair of integers p and q such that $p \geq 1$ and $(p, q) = 1$, choose a single solution p' of the congruence $pp' \equiv -1 \pmod{q}$, and then take for our transformations (7.12), all the transformations

$$(7.3) \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} \frac{pp' + 1}{q} & p' \\ p & q \end{pmatrix}.$$

We also must consider the poles in the x -plane due to the parabolic points. The point $\tau = i\infty$ corresponds to $x = 0$ while the other parabolic points correspond to points on the unit circle and hence are not included in $R(\infty)$. For $|x|$ small we have, by (1.3),

$$f(x) = \sum_{n=-\mu}^{\infty} a_n x^n,$$

and hence, since $y \neq 0$, we find, at $x = 0$,

$$\begin{aligned} \operatorname{Res} \frac{f(x)}{x-y} &= -\operatorname{Res} \frac{f(x)}{y} \frac{1}{1-x/y} = -\operatorname{Res} \frac{1}{y} \sum_{n=-\mu}^{\infty} a_n x^n \sum_{m=0}^{\infty} \left(\frac{x}{y}\right)^m \\ &= -\sum_{m=1}^{\infty} a_{-m} y^{-m}. \end{aligned}$$

From this, (7.22), (7.3), and the fact that y is distinct from the singularities of $f(x)$, we find that the sum of the residues of $f(x)/(x-y)$ at all the poles of $f(x)$ within the unit circle has the value

$$\begin{aligned} & -\sum_{m=1}^{\mu} a_{-m} y^{-m} + 2\pi i R e^{2\pi i(1-a)\sigma} (e^{2\pi i\sigma} - y)^{-1} \\ (7.4) \quad & -2\pi i R \sum_{p=1}^{\infty} \sum_{\substack{q=-\infty \\ (p,q)=1}}^{\infty} \epsilon \left(-q, p', p, -\frac{pp' + 1}{q} \right)^{-1} \\ & \quad \times (-i(p\sigma + q))^{-r-2} e^{2\pi i(1-a)\sigma'} (e^{2\pi i\sigma'} - y)^{-1}, \end{aligned}$$

where, in an effort to simplify the notation, we are using the abbreviation

$$(7.5) \quad \sigma' = \frac{((pp' + 1)/q)\sigma + p'}{p\sigma + q},$$

so σ' is a function of the indices of summation as well as of σ .

The infinite series (7.4) is absolutely convergent. To see this we first note that, since σ is in the fundamental region, there are only a finite number of its images of the type (7.5) which lie above any fixed line parallel to the real axis and above it. Also we have

$$\alpha < 1, \quad \Im(\sigma') > 0, \quad \left| \epsilon \left(-q, p', p, -\frac{pp' + 1}{q} \right) \right| = 1.$$

If y is in a region $0 < y_0 \leq |y| < 1$ then, leaving aside the finite number of terms for which

$$|e^{2\pi i \sigma'}| \leq \frac{1 + y_0}{2},$$

we see that the infinite series in (7.4) is majorized by the series

$$\frac{4\pi |R|}{1 - y_0} \sum_p \sum_q |p\sigma + q|^{-r-2},$$

which is known to converge for $r > 0$.

Since (7.4) is absolutely convergent we can rearrange its terms to agree with our definition of $R(\infty)$. Then we have:

THEOREM 2. *If the only singularities of the modular form $F(\tau)$ of Theorem 1, in the fundamental region, are a possible polar singularity at $i\infty$ and a simple pole with residue R at the point σ in the interior of the fundamental region, then the quantity $R(\infty)$ of Theorem 1 has the value (7.4) which is absolutely convergent for any y inside the unit circle which is not a singularity of $f(x)$.*

If $F(\tau)$ has several simple poles the value of $R(\infty)$ is the sum of corresponding expressions of the type (7.4). If $F(\tau)$ has a simple pole at a vertex of the fundamental region the restrictions on p and q in (7.4) have to be strengthened.

8. As in the case in which $F(\tau)$ is analytic in the upper half-plane,⁸ we can characterize the class of all functions which satisfy the conditions of Theorem 1. We shall find this characterization in a somewhat different way, basing it on a formula connecting the number of zeros and poles of a modular form. We suppose that $F(\tau)$ has Z zeros and P poles in the fundamental region omitting the vertices, $i\infty$, i , $\rho = e^{\pi i/3}$, and ρ^2 . Also we suppose that, at $i\infty$, $F(\tau)$ has an expansion of the form (1.3) with the associated constants α and μ , while it has zeros of orders s and t at i and ρ respectively. The integers

⁸ *Loc. cit.*, footnote 2, section 8.

s and t are to be taken negative if $F(\tau)$ has a pole at the corresponding points. Then these constants are connected by the formula

$$(8.11) \quad P + \mu - \alpha - \left(Z + \frac{s}{2} + \frac{t}{3} \right) = \frac{r}{12}.$$

This formula can be found by considering the integral,

$$(8.12) \quad \frac{1}{2\pi i} \int d(\log F(\tau)) = Z - P,$$

taken over the path formed by the sides of the fundamental region and a line parallel to the real axis and sufficiently far above it. Circular detours are made around the points i , ρ , ρ^2 , and any poles of $F(\tau)$; and then (8.11) is found as the limit of (8.12) as the horizontal line approaches $i\infty$ and the circular detours shrink down to their points.

We now multiply $F(\tau)$ by factors to obtain a new function which has no zeros or poles in the fundamental region, including the vertices. To cancel the Z zeros and P poles, which we may suppose to be at the points $\rho_1, \rho_2, \dots, \rho_Z$ and $\sigma_1, \sigma_2, \dots, \sigma_P$, respectively, we use the factor

$$(8.21) \quad \Theta(\tau) = \prod_{j=1}^P (J(\tau) - J(\sigma_j)) \prod_{k=1}^Z (J(\tau) - J(\rho_k))^{-1},$$

which may introduce new zeros or poles at $i\infty$ but nowhere else. To take care of the finite vertices we use the factor⁹

$$(8.22) \quad \Phi(\tau) = \left(\sqrt{J(\tau) - 1} \right)^{-s} (\sqrt[3]{J(\tau)})^{-t}.$$

Finally the factor $\eta(\tau)^{2r}$ will suffice for the point $\tau = i\infty$. To see this we expand the function

$$(8.23) \quad \Psi(\tau) = F(\tau) \eta(\tau)^{2r} \Phi(\tau) \Theta(\tau)$$

about the point $i\infty$, using the known expansions of $J(\tau)$ and $\eta(\tau)$, and the expansion (1.3) of $F(\tau)$. We then have

$$\Psi(\tau) = \exp\left\{2\pi i\tau\left(\alpha - \mu + \frac{r}{12} + \frac{s}{2} + \frac{t}{3} - P + Z\right)\right\} \sum_{n=0}^{\infty} c_n e^{2\pi i n \tau} = \sum_{n=0}^{\infty} c_n e^{2\pi i n \tau},$$

where we have used (8.11) to obtain the last equality.

The form $\Psi(\tau)$ is a modular function. To prove this we need consider only the two generators of the modular group. For the function (8.21) we have

⁹ We use the usual determinations of the roots, *loc. cit.*, footnote 2, p. 449.

$$(8.31) \quad \Theta(\tau+1) = \Theta(\tau), \quad \Theta\left(\frac{-1}{\tau}\right) = \Theta(\tau),$$

since $J(\tau)$ is a modular function. The function (8.22) has the transformation formula¹⁰

$$(8.32) \quad \Phi(\tau+1) = \exp\left\{2\pi i\left(\frac{s}{2} + \frac{t}{3}\right)\right\} \Phi(\tau), \quad \Phi\left(\frac{-1}{\tau}\right) = e^{\pi i s} \Phi(\tau),$$

and $\eta(\tau)^{2r}$ the formula

$$(8.33) \quad \eta(\tau+1)^{2r} = \exp\left(2\pi i \frac{r}{12}\right) \eta(\tau)^{2r}, \quad \eta\left(\frac{-1}{\tau}\right)^{2r} = (-i\tau)^r \eta(\tau)^{2r}.$$

From (8.23), (1.12), (8.31), (8.32), and (8.33) we then find

$$(8.41) \quad \Psi(\tau+1) = \exp\left\{2\pi i\left(\alpha + \frac{r}{12} + \frac{s}{2} + \frac{t}{3}\right)\right\} \Psi(\tau),$$

but from (8.11) we have

$$\alpha + \frac{r}{12} + \frac{s}{2} + \frac{t}{3} = P + \mu - Z,$$

an integer, and hence (8.41) reduces to

$$(8.42) \quad \Psi(\tau+1) = \Psi(\tau).$$

Also, from (8.23), (1.11), (8.31), (8.32), and (8.33), we get

$$(8.43) \quad \Psi\left(\frac{-1}{\tau}\right) = \epsilon_0 e^{\pi i s} \Psi(\tau),$$

where

$$(8.44) \quad \epsilon_0 = \epsilon(0, -1, 1, 0).$$

Now the value of ϵ_0 is known¹¹ to be

$$(8.45) \quad \epsilon_0 = \exp\left\{2\pi i\left(-3\alpha - \frac{r}{4}\right)\right\},$$

and therefore we have, using (8.11),

$$\epsilon_0 e^{\pi i s} = \exp\left\{2\pi i\left(-3\alpha - \frac{r}{4} - \frac{3s}{2}\right)\right\} = \exp\{2\pi i(-3P - 3\mu + 3Z + t)\} = 1.$$

and hence (8.43) reduces to

$$(8.46) \quad \Psi\left(\frac{-1}{\tau}\right) = \Psi(\tau).$$

¹⁰ *Loc. cit.*, footnote 2, p. 449, formulas (8.63) and (8.64).

¹¹ *Loc. cit.*, footnote 2, p. 445, formula (6.7). The proof used there applies without change to our case in which $F(\tau)$ has poles in the fundamental region.

Formulas (8.42) and (8.46) show that $\Psi(\tau)$ is a modular function. Since it has no singularities it is a constant, and, by (8.23), we have

$$(8.5) \quad F(\tau) = K\eta(\tau)^{-2r}\Phi(\tau)^{-1}\Theta(\tau)^{-1} \\ = K\eta(\tau)^{-2r} \prod_{j=1}^P (J(\tau) - J(\sigma_j))^{-1} \prod_{k=1}^Z (J(\tau) - J(\rho_k)) \\ \times \left(\sqrt{J(\tau) - 1} \right)^s (\sqrt[3]{J(\tau)})^t.$$

Any form $F(\tau)$ that satisfies the conditions of Theorem 1 can then be expressed in the form (8.5). Conversely, it is clear that any form (8.5) satisfies these conditions provided we have $r > 0$ and the σ_j and ρ_k distinct from the vertices of the fundamental region. The numbers μ and α associated with this form may be obtained from (8.11). Since μ is an integer and $0 \leq \alpha < 1$ we have

$$(8.61) \quad \mu = - \left[-\frac{r}{12} + P - Z - \frac{s}{2} - \frac{t}{3} \right],$$

and

$$(8.62) \quad \alpha = -\frac{r}{12} + P - Z - \frac{s}{2} - \frac{t}{3} - \left[-\frac{r}{12} + P - Z - \frac{s}{2} - \frac{t}{3} \right].$$

We can express the roots (8.22) in terms of $g_2(1, \tau)$, $g_3(1, \tau)$, and $\eta(\tau)$ and then write (8.5) in the form

$$(8.7) \quad F(\tau) = K\eta(\tau)^{-2r-12s-8t} \\ \times g_2(1, \tau)^t g_3(1, \tau)^s \prod_{j=1}^P (J(\tau) - J(\sigma_j))^{-1} \prod_{k=1}^Z (J(\tau) - J(\rho_k)),$$

where the constant K has changed its value.

9. The discussion of the $\epsilon(a, b, c, d)$, given in the paper¹² referred to above, for forms analytic in the upper half-plane, applies directly to our case also. Hence we can immediately state the

THEOREM 3. *The modular form, (8.7), of dimension r , satisfies the transformation equations*

$$(9.11) \quad F\left(\frac{a\tau + b}{c\tau + d}\right) = \epsilon(a, b, c, d) (-i(c\tau + d))^{-r} F(\tau), \quad c > 0,$$

and

$$(9.12) \quad F(\tau + 1) = e^{2\pi i a} F(\tau),$$

¹² *Loc. cit.*, footnote 2, section 9.

with the multiplier

$$(9.21) \quad \epsilon(a, b, c, d) = \exp \left\{ 2\pi i \left(r \cdot s(a, c) - r \frac{a+d}{12c} + \frac{s}{2} (b(a+d) + a(b-c) + bc) + \frac{t}{3} (a+d)(b-c)(ad+bc) \right) \right\},$$

and with the α of (8.62), where

$$(9.22) \quad s(a, c) = \sum_{l=1}^{c-1} \frac{l}{c} \left(\frac{al}{c} - \left[\frac{al}{c} \right] - \frac{1}{2} \right).$$

Conversely, all modular forms satisfying the conditions of Theorem 1 are contained in (8.7).

We can apply Theorem 1 to the form (8.7) and evaluate the $A_{k,v}(n)$ by means of (9.21) to obtain

THEOREM 4. If $r > 0$, the modular form (8.7), in which the ρ_k and σ_j are distinct from the vertices of the fundamental region, has the expansion

$$(9.31) \quad f(y) = \sum_{k=1}^{\infty} \sum_{v=1}^{\mu} a_{-v} \frac{1}{k} \sum_{n=0}^{\infty} A_{k,v}(n) \left(\frac{v-\alpha}{n+\alpha} \right)^{(r+1)/2} \times I_{r+1} \left(\frac{4\pi}{k} \sqrt{(v-\alpha)(n+\alpha)} \right) y^n - R(\infty),$$

where α and μ are given by (8.61) and (8.62); the a_{-v} are the coefficients of $e^{-2\pi i v \tau}$ in the expansion of $e^{-2\pi i a \tau} F(\tau)$, valid for $\Im(\tau)$ sufficiently large; $R(\infty)$ is the sum of the residues of $f(x)/(x-y)$ at the poles of $f(x)$ within the unit circle, as described in section 6; and where

$$(9.32) \quad A_{k,v}(n) = \sum_{\substack{0 \leq h < k \\ (h,k)=1}} \omega_r(h, k) \xi_1(h, k) {}^s\xi_2(h, k)^t \times \exp \left\{ -\frac{2\pi i}{k} ((v-\mu-P+Z)h' + (n+\mu+P-Z)h) \right\},$$

with ¹³

$$(9.33) \quad \begin{aligned} \omega_r(h, k) &= \exp\{2\pi i r \cdot s(h, k)\}, \\ \xi_1(h, k) &= \exp \left\{ \pi i \left(\frac{h(h'^2+1)}{k} - hk - 1 - hh' \right) \right\} \\ &= \exp \left\{ \pi i \left(-\frac{h'(h^2+1)}{k} + h'k + hh' + 1 \right) \right\}, \\ \xi_2(h, k) &= \exp \left\{ \frac{2\pi i}{3} (h-h') \left(\left(\frac{hh'+1}{k} + k \right) (2hh'+1) + \frac{1}{k} \right) \right\}. \end{aligned}$$

¹³ These are the same as the functions (9.54) of the paper referred to in footnote 2. It should be noted that there is an error in the expression for $\xi_2(h, k)$ in that paper.

10. The modular form

$$(10.11) \quad F_r(\tau) = \frac{1}{1728} \eta(\tau)^{-2r} (J(\tau) - J(\sigma))^{-1}$$

affords an interesting example. We limit r to be positive and σ to be distinct from the vertices of the fundamental region. Then $F_r(\tau)$ is of the form (8.7) with

$$(10.12) \quad s = t = Z = 0, \quad P = 1,$$

and (8.61) and (8.62) yield the values

$$(10.13) \quad \mu = -\left[-\frac{r}{12} + 1\right], \quad \alpha = -\frac{r}{12} + 1 - \left[-\frac{r}{12} + 1\right].$$

Corresponding to (1.3) we have the expansion

$$(10.14) \quad e^{-2\pi i a \tau} F_r(\tau) = x^{r/12 - 1 - \mu} \cdot x^{-(r/12)} (1 + x + 2x^2 + \dots)^{2r} \\ \times (x^{-1} + (744 - 1728J(\sigma)) + 196884x + \dots)^{-1} \\ = x^{-1-\mu} (x - (744 - 2r - 1728J(\sigma))x^2 + \dots) \\ = x^{-\mu} - (744 - 2r - 1728J(\sigma))x^{-\mu+1} + \dots,$$

and, therefore, the values

$$(10.15) \quad a_{-\mu} = 1, \quad a_{-\mu+1} = 1728J(\sigma) - 744 + 2r, \dots$$

If $0 < r \leq 12$ we have $\mu = 0$ and hence, by Theorem 4,

$$(10.2) \quad F_r(\tau) = e^{2\pi i a \tau} f_r(e^{2\pi i \tau}), \quad f_r(y) = R(\infty).$$

The value of $R(\infty)$ may be found by Theorem 2. The residue of $F_r(\tau)$ at the point $\tau = \sigma$ is

$$(10.3) \quad R = \frac{1}{1728} \eta(\sigma)^{-2r} J'(\sigma)^{-1},$$

where $J'(\sigma)$ is the derivative of $J(\tau)$ at $\tau = \sigma$. The value of $\epsilon\left(-q, p', p, -\frac{pp' + 1}{q}\right)$ in (7.4) may be replaced by its value (9.21) and we then have

$$(10.41) \quad R(\infty) = \frac{2\pi i}{1728} \eta(\sigma)^{-2r} J'(\sigma)^{-1} \{e^{2\pi i(1-\alpha)\sigma} (e^{2\pi i \sigma} - y)^{-1} \\ - \sum_{p=1}^{\infty} \sum_{\substack{q=-\infty \\ (p,q)=1}}^{\infty} \xi(p, q)^r (-i(p\sigma + q))^{-r-2} e^{2\pi i(1-\alpha)\sigma'} (e^{2\pi i \sigma'} - y)^{-1}\},$$

with the α of (10.13) and where we have used the abbreviations (7.5) and

$$(10.42) \quad \zeta(p, q) = \exp \left\{ -2\pi i \left(s(-q, p) - \frac{q + (pp' + 1)/q}{12p} \right) \right\}.$$

With this value for $R(\infty)$ we then have the desired expansion of $F_r(\tau)$ in (10.2).

A formula concerning modular forms of negative dimension can be obtained from our expansions of $F_r(\tau)$. For $\Im(\tau) > \Im(\sigma)$ we expand each summand of (10.41) in a geometric series and then have, using (10.2)

$$F_r(\tau) = -\frac{2\pi i}{1728} \eta(\sigma)^{-2r} J'(\sigma)^{-1} e^{2\pi i a \tau} \left\{ e^{-2\pi i a \sigma} \sum_{n=0}^{\infty} e^{2\pi i n(\tau - \sigma)} - \sum_{p=1}^{\infty} \sum_{\substack{q=-\infty \\ (p,q)=1}}^{\infty} \zeta(p, q)^r (-i(p\sigma + q))^{-r-2} e^{-2\pi i a \sigma'} \sum_{n=0}^{\infty} e^{2\pi i n(\tau - \sigma')} \right\}.$$

On the other hand, by (10.14) with the values (10.13), we also have

$$F_r(\tau) = e^{2\pi i a \tau} \{ 1 - (744 - 2r - 1728J(\sigma)) e^{2\pi i \tau} + \dots \},$$

valid for $\Im(\tau) > \Im(\sigma)$. A comparison of the leading terms now yields the result

$$(10.51) \quad 1728\eta(\sigma)^{2r} J'(\sigma) = -2\pi i \left\{ e^{-2\pi i a \sigma} - \sum_{p=1}^{\infty} \sum_{\substack{q=-\infty \\ (p,q)=1}}^{\infty} \zeta(p, q)^r (-i(p\sigma + q))^{-r-2} e^{-2\pi i a \sigma'} \right\},$$

for $0 < r \leq 12$ and σ inside the fundamental region. By analytic continuation the restriction on σ can be removed and we have (10.51) for $\Im(\sigma) > 0$.

For the case $r=12$ this formula simplifies considerably. Making use of the result ¹⁴

$$(10.52) \quad 12s(h, k) \equiv \frac{h - h'}{k} \pmod{1},$$

for $hh' \equiv -1 \pmod{k}$, we have, since $-q \cdot \frac{pp' + 1}{q} \equiv -1 \pmod{p}$,

$$12s(-q, p) \equiv \frac{-q - (pp' + 1)/q}{p} \pmod{1},$$

and therefore

$$(10.53) \quad \zeta(p, q)^{12} = 1.$$

¹⁴ H. Rademacher, "Zur Theorie der Modulfunktionen," *Crelle*, vol. 167 (1931), pp. 312-336, in particular p. 321, formula (2.51).

Then (10.51) reduces to the expression

$$1728\eta(\sigma)^{24}J'(\sigma) = -2\pi i \left\{ 1 + \sum_{p=1}^{\infty} \sum_{\substack{q=-\infty \\ (p,q)=1}}^{\infty} (p\sigma + q)^{-14} \right\},$$

which can easily be reduced to the more usual form

$$1728\eta(\sigma)^{24}J'(\sigma) = -\frac{3 \cdot 13!}{2^{11}\pi^{13}} i \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} (p\sigma + q)^{-14}.$$

Returning to our original example, (10.11), we now consider the case $12 < r \leq 24$. We now have $\mu = 1$ and hence, by Theorem 4 with the values (10.13) and (10.15), we find

$$(10.61) \quad \begin{aligned} F_r(\tau) &= e^{2\pi i \alpha \tau} f_r(e^{2\pi i \tau}), \\ f_r(y) &= 2\pi \sum_{k=1}^{\infty} \frac{1}{k} \sum_{h=0}^{\infty} A_{k,1}^{(r)}(n) \left(\frac{1-\alpha}{n+\alpha} \right)^{(r+1)/2} \\ &\quad \times I_{r+1} \left(\frac{4\pi}{k} \sqrt{(1-\alpha)(n+\alpha)} \right) y^n - R(\infty), \end{aligned}$$

where $A_{k,1}^{(r)}(n)$ has the value

$$(10.62) \quad A_{k,1}^{(r)}(n) = \sum_{\substack{0 \leq h < k \\ (h,k)=1}} e^{2\pi i r \cdot s(h,k)} \exp \left\{ -\frac{2\pi i}{k} (-h' + (n+2)h) \right\}.$$

The value of $R(\infty)$ may again be found by Theorem 2. It is found to have the value (10.41) with the added term $-y^{-1}$.

For $r = 24$ the expression for $F_r(\tau)$ again simplifies. In this case we have $\alpha = 0$ and hence, making use of (10.53), we find

$$(10.71) \quad \begin{aligned} F_{24}(\tau) &= 2\pi \sum_{k=1}^{\infty} \frac{1}{k} \sum_{n=0}^{\infty} A_{k,1}^{(24)}(n) n^{-(25/2)} J_{25} \left(\frac{4\pi}{k} \sqrt{n} \right) e^{2\pi i n \tau} + e^{-2\pi i \tau} \\ &\quad - \frac{2\pi i}{1728} \eta(\sigma)^{-48} J'(\sigma)^{-1} \left\{ (1 - e^{2\pi i (\tau-\sigma)})^{-1} \right. \\ &\quad \left. + \sum_{p=1}^{\infty} \sum_{\substack{q=-\infty \\ (p,q)=1}}^{\infty} (p\sigma + q)^{-26} (1 - e^{2\pi i (\tau-\sigma')})^{-1} \right\}, \end{aligned}$$

where, by (10.62) and (10.52),

$$(10.72) \quad A_{k,1}^{(24)}(n) = \sum_{\substack{0 \leq h < k \\ (h,k)=1}} \exp \left\{ -\frac{2\pi i}{k} (nh + h') \right\}.$$

It is of interest to ask whether the two parts of our expansions are each separately modular forms. We shall not answer this question completely but shall merely give some indications as to the answer by considering the particular example (10.71). We first write $F_{24}(\tau) = G(\tau) + H(\tau)$ where $G(\tau)$ is the part of (10.71) which is a power series in $e^{2\pi i \tau}$ and $H(\tau)$ is the

remainder. The series $G(\tau)$ converges for all τ in the upper half-plane so $G(\tau)$ has no finite poles. If $G(\tau)$ is a modular form of positive dimension belonging to the full group then it can be seen¹⁵ from its expansion in (10.71) that it is of dimension 24. The characteristic constants of $G(\tau)$ then have the values $r=24$, $\alpha=0$, $\mu=1$, $P=0$, $Z \geq 0$, $s \geq 0$ and $t \geq 0$. However these values contradict equation (8.11) and therefore $G(\tau)$ is not a modular form of positive dimension belonging to the full group.

We now turn to the function $H(\tau)$. From its value in (10.71) we have $H(\tau+1) = H(\tau)$ and hence $\alpha=0$. Also $H(\tau)$ remains finite as $\tau \rightarrow i\infty$ and therefore we have $\mu \leq 0$. Since $P=1$, $Z \geq 0$, $s \geq 0$ and $t \geq 0$ we see, by (8.11), that if $H(\tau)$ is a modular form of positive dimension r belonging to the full group, then we have $\mu=0$, $Z=0$ and

$$(10.81) \quad \frac{r}{12} = 1 - \frac{s}{2} - \frac{t}{3}.$$

Then by Theorem 3, $H(\tau)$ is one of the forms

$$(10.82) \quad H(\tau) = K\eta(\tau)^{-2r}(J(\tau) - J(\sigma))^{-1}(\sqrt{J(\tau) - 1})^s(\sqrt[3]{J(\tau)})^t,$$

where K may depend on σ . From our definitions of the functions we now have

$$H(\tau) = \frac{1}{1728} \eta(\tau)^{-48}(J(\tau) - J(\sigma))^{-1} - G(\tau),$$

which we combine with (10.81) to obtain

$$(10.83) \quad \frac{\eta(\tau)^{-48}}{1728} - G(\tau)(J(\tau) - J(\sigma))^{-1} = K\eta(\tau)^{-2r}(\sqrt{J(\tau) - 1})^s(\sqrt[3]{J(\tau)})^t.$$

This equation must be valid for $\tau \neq \sigma$ and σ within the fundamental region. By continuation it then holds for all τ and σ in the upper half-plane. From (10.83) we see that K is a modular form of dimension 0 in the variable σ . However if we set $\tau = \sigma$ in (10.82) we see that K is of dimension $(24-r)$. Therefore we have $r=24$. This contradicts the condition (10.81) and we therefore see that $H(\tau)$ is not a modular form of positive dimension belonging to the full group.

UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA, PENNSYLVANIA.

¹⁵ *Loc. cit.*, footnote 2, p. 458, footnote 18.

THE EXPONENTIAL REPRESENTATION OF AUTOMORPHS OF A SYMMETRIC OR HERMITIAN MATRIX.*

By JOHN WILLIAMSON.

In a previous paper¹ the exponential representation of canonical matrices was studied. These canonical matrices are automorphs of the normal form of a skew symmetric matrix. The corresponding problem, when the skew symmetric matrix is replaced by a symmetric or hermitian matrix, is considered here. Three cases are treated: when the matrix is symmetric over the complex field, when the matrix is hermitian, and when the matrix is symmetric over the real field. Of these three the last is by far the most interesting. The methods employed are similar to those of the paper quoted above and, as in many cases the proofs are practically identical, they will not always be given in detail.

1. We shall consider square matrices over the real or the complex field and, if H is such a matrix, we shall mean by H^* either the transposed or else the conjugate transposed of H . Let H be non-singular and let $H = H^*$, so that H is either symmetric or hermitian. If $G = -G^*$ and

$$(1) \quad C = \exp(HG),$$

then

$$(2) \quad CHC^* = H;$$

$$\text{for } CHC^* = \exp(HG)H \exp(HG)^* = \exp(HG)H \exp(G^*H^*) \\ = \exp(HG)H \exp(-GH) = \exp(HG) \exp(-HG)H = H.$$

Since the determinant of a matrix is the product of the latent roots of the matrix,

$$|\exp(HG)| = \exp(\text{trace of } HG).$$

If

$$H = (h_{ij}) \text{ and } G = (g_{ij}), \quad (i, j = 1, 2, \dots, n), \\ t = \text{trace } (HG) = \sum_{i=1}^n \sum_{j=1}^n h_{ij}g_{ji} = - \sum_{i=1}^n \sum_{j=1}^n \bar{h}_{ji}\bar{g}_{ij} = -\bar{t}.$$

Therefore, when $*$ denotes conjugate transposed, the trace of HG is a pure imaginary number and the determinant of $\exp(HG)$ has absolute value one.

* Received April 20, 1939.

¹ John Williamson, "The exponential representation of canonical matrices," *American Journal of Mathematics*, vol. 61 (1939), pp. 897-911. This paper will be referred to as I.

In the other case, when $*$ denotes transposed, $t = \bar{t}$ and therefore t is zero. Consequently,

$$(3) \quad |\exp(HG)| = \exp(0) = +1.$$

We shall determine here necessary and sufficient conditions, that a matrix C , which satisfies (2), shall have an exponential representation of the form (1). If a matrix C satisfies (2), since H is non-singular, $|C| |C|^* = 1$. Therefore the absolute value of $|C|$ is unity and, when $*$ denotes transposed, $|C| = \pm 1$. As a consequence of (3), when H is symmetric, we must restrict our consideration to those matrices C , whose determinants have the value plus one.

$$\text{If } A = HG,$$

$$AH = HGH = -HA^* = Hf(A^*),$$

where $f(x) \equiv -x$. Hence A is normal² with respect to H .

Let P be a non-singular matrix and let

$$(4) \quad PHP^* = H_1 \text{ and } PCP^{-1} = C_1.$$

Then, if $CHC^* = \bar{H}$, $C_1H_1C_1^* = H_1$. Similarly, if $C = \exp A = \exp(HG)$, $C_1 = \exp A_1 = \exp(H_1G_1)$, where $G_1 = -G^*$.

Since

$$(5) \quad A_1 = PAP^{-1},$$

A_1 is normal with respect to H_1 . Further the matrix C is also normal with respect to H and the matrix C_1 normal with respect to H_1 , where the defining polynomial is $f(x) \equiv x^{-1}$. For brevity, when equations (4) are satisfied for some matrix P , we shall write $(H, C) \sim (H_1, C_1)$ and, when (4) and (5) are satisfied, $(H, A) \approx (H_1, A_1)$. Since both the symbols \sim and \approx have all the properties of an equivalence relation we have,

RESULT (a). *A matrix C , which satisfies (2), has an exponential representation of the form (1), if, and only if, there exists a pair $(H_1, C_1) \sim (H, C)$ and a pair $(H_1, A_1) \approx (H, A)$, where $C_1 = \exp A_1$.*

Since canonical pairs $(H_1, A_1) \approx (H, A)$ and canonical pairs $(H_1, C_1) \sim (H, C)$ are known, it is only necessary³ to compare the matrices $\exp A_1$ with the known matrices C_1 .

² John Williamson, "Matrices normal with respect to an hermitian matrix," *American Journal of Mathematics*, vol. 60 (April, 1938), pp. 355-373; "Normal matrices over an arbitrary field of characteristic zero," *American Journal of Mathematics*, vol. 61 (April, 1939). These papers will be referred to as II and III respectively.

³ Cf. I, page 898.

2. Hermitian case. Let H be a non-singular hermitian matrix over the complex field and let H^* denote the conjugate transposed of H . Since C is normal with respect to H and $f(x) \equiv x^{-1}$, the canonical forms $(H_1, C_1) \sim (H, C)$ can easily be written down as particular cases from the general results of II or III. The actual forms of H_1 and C_1 are, however, not necessary. It is sufficient for our purposes, that the matrices H_1 and C_1 are similarly partitioned diagonal block matrices. These blocks depend, though not always uniquely, on the elementary divisors of $C - \lambda E$ or, as we shall say, on the elementary divisors of C . These elementary divisors are of two distinct types;⁴

Type (i): the pair $(\lambda - a)^r, (\lambda - a^{-1})^r$, where $|a| \neq 1$, and

Type (ii): $(\lambda - a)^r$, where $|a| = 1$.

Since the matrices of the canonical pair are similarly partitioned diagonal block matrices, the general case reduces to the consideration of two particular cases; that, in which C has a single pair of elementary divisors of type (i), and that, in which C has a single elementary divisor of type (ii). In the first of these cases the canonical pair $(H_1, C_1) \sim (H, C)$ is unique; in the second there are two non-equivalent pairs $(\rho X_1, C_1)$, where C_1 and X_1 are unique but $\rho = \pm 1$.

The matrices of the canonical pair $(H_1, A_1) \approx (H, A)$ are again similarly partitioned diagonal block matrices depending on the elementary divisors of A . These elementary divisors are of two distinct types;

Type (α): the pair $(\lambda - p)^r, (\lambda + \bar{p})^r$, where $p \neq -\bar{p}$, and

Type (β): $(\lambda - p)^r$, where $p = -\bar{p}$.

If A has the single pair of elementary divisors of type (α), A_1 and H_1 are unique. The matrix $\exp A_1$ has the single pair of elementary divisors $(\lambda - a)^r, (\lambda - a^{-1})^r$, where $a = \exp p$. Since $p \neq -\bar{p}$, $|a| \neq 1$. Further, if $|a| \neq 1$, we can always determine $p = \log a$ where $p \neq -\bar{p}$. Consequently every matrix C with the single pair of elementary divisors of type (i) has an exponential representation of the form (1). If A has the single elementary divisor of type (β), the canonical pair $(H_1, A_1) \approx (H, A)$ is not unique. The matrix A_1 is unique but $H_1 = \rho Y_1$, where $\rho = \pm 1$ and Y_1 is unique.⁵ The matrix $\exp A_1$ has the single elementary divisor $(\lambda - a)^r$, where $a = \exp p$. Since $p = -\bar{p}$, $|a| = 1$. Conversely, if $|a| = 1$, $p = \log a$ satisfies $p = -\bar{p}$.

⁴ II, page 360; John Williamson, "Quasi-unitary matrices," *Duke Mathematical Journal*, vol. 3 (December, 1937), no. 4, page 414.

⁵ The matrix Y_1 may be taken to be the same as the matrix X_1 in the canonical pair H_1, C_1 of type (ii).

Since there are two distinct canonical pairs Y_1, A_1 and $-Y_1, A_1$, we get two distinct canonical pairs $\pm Y_1, \exp A_1$ for the pair H, C , where C has a single elementary divisor of type (ii). Therefore we have the theorem:

THEOREM 1. *If C is a conjunctive automorph of the non-singular hermitian matrix H , $C = \exp(HG)$, where G is anti-hermitian.⁶*

If $H = E$, the identity matrix, C is unitary and we obtain the well known corollary:

COROLLARY 1. *Every unitary matrix U may be written in the form $U = \exp G$, where G is anti-hermitian.⁷*

Since, if H is hermitian, iH is anti-hermitian we also have the corollary:

COROLLARY 2. *If C is a conjunctive automorph of the non-singular anti-hermitian matrix H , $C = \exp HG$, where G is hermitian.*

3. Complex field. Let H be a non-singular symmetric matrix over the complex field and let $*$ denote transposed. The canonical pairs H_1, C_1 and H_1, A_1 are now uniquely determined by the elementary divisors of C and A respectively. As in § 2 the general case of a matrix C can be deduced from two particular cases; that, in which C has a single pair of elementary divisors $(\lambda - a)^r, (\lambda - a^{-1})^r$, and that, in which C has a single elementary divisor $(\lambda \pm 1)^{2k+1}$. The two particular cases, from which the general case of a matrix A may be deduced, are those, in which A has a single pair of elementary divisors $(\lambda - p)^r, (\lambda + p)^r$ and a single elementary divisor λ^{2k+1} . It follows immediately that every matrix C with a single pair of elementary divisors $(\lambda - a)^r, (\lambda - a^{-1})^r$ does have an exponential representation of the form (1), as does a matrix C with the single elementary divisor $(\lambda - 1)^{2k+1}$. On the other hand a matrix C with a single elementary divisor $(\lambda + 1)^{2k+1}$ does not have⁸ an exponential representation of the form (1). We therefore have the theorem,

THEOREM 2. *Let C be an automorph of the non-singular symmetric matrix H over the complex field. The matrix C has an exponential representation of the form $C = \exp HG$, with a skew symmetric G , if, and only if,*

⁶ This answers for the case of finite matrices a question raised by Aurel Wintner, "Über die automorphen Transformationen beschränkter nicht-singulärer hermitescher Formen," *Mathematische Zeitschrift*, vol. 39 (1933), page 263.

⁷ Aurel Wintner, "Spektraltheorie der unendlichen Matrizen," (Leipzig, 1929), page 217.

⁸ This is obvious directly, since $|C| = -1$.

no elementary divisor $(\lambda + 1)^{2k+1}$ occurs an odd number of times among the elementary divisors of C .

Let C have the single elementary divisor $(\lambda + 1)^r$, where $r = 2k + 1$. Let E_r , U_r and T_r be respectively the unit matrix, the auxiliary unit matrix and the counter unit matrix of order r . Then, if ⁹

$$(6) \quad X_r = [1, -1, \dots, (-1)^{r+1}]T_r,$$

we may take $H_1 = X_r$ and $C_1 = \Gamma_r$, where

$$(7) \quad \Gamma_r = -\exp U_r.$$

If B_r is the diagonal block matrix given by $B_r = [bE_k, 1, b^{-1}E_k]$,

$$B_r X_r B_r^* = X_r,$$

and therefore, if

$$(8) \quad D_r = B_r \Gamma_r,$$

D_r is an automorph of X_r . The elementary divisors of D_r are obviously $(\lambda + b)^k$, $(\lambda + b^{-1})^k$ and $(\lambda + 1)$. Further

$$(9) \quad \lim_{b \rightarrow 1} D_r = \Gamma_r.$$

If C has only the elementary divisors $(\lambda + 1)^{r_i}$, where $r_i = 2k_i + 1$, $i = 1, 2, \dots, s$, we may take

$$(10) \quad C_1 = [\Gamma_{r_1}, \Gamma_{r_2}, \dots, \Gamma_{r_s}] \quad \text{and} \quad H_1 = [X_{r_1}, X_{r_2}, \dots, X_{r_s}],$$

where Γ_r and X_r are defined by (6) and (7). If

$$(11) \quad D = [D_{r_1}, D_{r_2}, \dots, D_{r_s}], \quad \text{where } D_r \text{ is defined by (8),}$$

then, as a consequence of (9),

$$\lim_{b \rightarrow 1} D = C_1.$$

The elementary divisors of D are the s pairs $(\lambda + b)^{r_i}$, $(\lambda + b^{-1})^{r_i}$ and the s elementary divisors $(\lambda + 1)$. If s is even, D has, as a consequence of Theorem 2, an exponential representation $\exp(H_1 G_1)$. Now, if C is a proper automorph of H , so that $|C|$ is ± 1 , and C does not have an exponential representation of the form (1), the matrices H_1, C_1 of the canonical pair must include as submatrices the matrices given by (10), where s is even. If, in C_1 , this submatrix be replaced by the matrix D in (11), the resulting matrix has an exponential representation $\exp(H_1 G_1)$ and its limit as $b \rightarrow 1$ is C_1 . Therefore we have the theorem:

⁹ See II, page 356, or III, page 337, or I, page 903, footnote 13.

THEOREM 3. *If C is a proper automorph of the non-singular symmetric matrix H over the complex field, C is either of the form $\exp(HG)$, where G is skew symmetric or is the limit of automorphs, which are.*

It is of course obvious that, if $|C| = -1$, $C \neq \text{Lim } D$, where $D = \exp(HG)$.

If C is not representable in the form (1), we may write $C_1 = [C_2, C_3]$, where all the elementary divisors of C_3 are of the form $(\lambda + 1)^r$ while none of C_2 are of that form. If E_i is the unit matrix of the same order as C_i , the matrix $J = [E_2, -E_3]$ is an automorph of H_1 and is of period two. The matrix JC_1 is an automorph of H_1 and, since no latent root of JC_1 has the value minus one, $JC_1 = \exp(H_1G_1)$. Accordingly we have the theorem;

THEOREM 4. *If C is an automorph of H , which does not have an exponential representation of the form (1), there exists an automorph D of H , such that DC does have such a representation. The automorph D is of period two.*

If C is proper, the number of elementary divisors of D of the form $\lambda + 1$ is always even and we therefore have

COROLLARY 1. *If C is proper, the matrix D of Theorem 4 has an exponential representation of the form (1).*

4. The real field. Let H be a non-singular real symmetric matrix and let * denote transposed. The general canonical pair $(H_1, C_1) \sim (H, C)$ can again be deduced from that of several simple types of the matrix C . The simple matrix C has elementary divisors of the following types:

Type (i): a single pair of real elementary divisors $(\lambda - a)^r, (\lambda - a^{-1})^r$.

Type (ii): the four elementary divisors

$$(\lambda - a)^r, (\lambda - \bar{a})^r, (\lambda - a^{-1})^r, (\lambda - \bar{a}^{-1})^r; \quad |a| \neq 1, a \neq \bar{a}.$$

Type (iii): a single pair of elementary divisors

$$(\lambda - a)^r, (\lambda - \bar{a})^r; \quad |a| = 1, a \neq \bar{a}.$$

Type (iv): the single elementary divisor $(\lambda \pm 1)^{2k+1}$.

In types (i) and (ii) H_1 and C_1 are unique but in types (iii) and (iv), while C_1 is unique, $H_1 = \rho X_1$, where $\rho = \pm 1$ and X_1 is unique.¹⁰

The general canonical pair $(H_1, A_1) \approx (H, A)$ can also be deduced from

¹⁰ II, page 371, or III, page 351.

that of several simple types of the matrix A . The simple matrix A has elementary divisors of the following types:

Type (α): a single pair of real elementary divisors

$$(\lambda - p)^r, (\lambda + p)^r.$$

Type (β): the four elementary divisors

$$(\lambda - p)^r, (\lambda - \bar{p})^r, (\lambda + p)^r, (\lambda + \bar{p})^r; \quad p \neq \bar{p}, p \neq -\bar{p}.$$

Type (γ): a single pair of elementary divisors

$$(\lambda - p)^r, (\lambda - \bar{p})^r; \quad p = -\bar{p}.$$

Type (δ): the single elementary divisor λ^{2k+1} .

In types (α) and (β) the matrices H_1 and A_1 are unique, while in types (γ) and (δ), A_1 is unique but $H_1 = \rho Y_1$, where $\rho = \pm 1$. The canonical forms are of course all real.

If A is of type (α), $\exp A$ has the single pair of elementary divisors $(\lambda - a)^r, (\lambda - a^{-1})^r$, where, since p is real, $a = \exp p$ is positive. If A is of type (β), the matrix $\exp A$ has the four elementary divisors $(\lambda - a)^r, (\lambda - \bar{a})^r, (\lambda - a^{-1})^r, (\lambda - \bar{a}^{-1})^r$, where $a = \exp p$. Since the real part of p is not zero the absolute value of a is not one. The number a is real if, and only if, the imaginary part of p is an integral multiple of π , in which case $a = \bar{a}$ and $a^{-1} = \bar{a}^{-1}$. Therefore, if A is of type (β), $\exp A$ is a matrix C of type (ii) or else a matrix C with exactly two equal pairs of elementary divisors of type (i). In this last case it should be noted that a may either be positive or negative.¹¹

If A is of type (γ), the matrix $\exp A$ has the single pair of elementary divisors $(\lambda - a)^r, (\lambda - a^{-1})^r$, where $a = \exp p$ and, since $p = -\bar{p}$, $|a| = 1$. Further, if $|a| = 1$, and a is not real, we can always determine p to satisfy both of the equations $a = \exp p$ and $p = -\bar{p}$. Since $H_1 = \rho Y_1$, $\rho = \pm 1$, the matrices H , $\exp A$ have two distinct canonical forms. If the imaginary part of p is an integral multiple of π , $a = \pm 1$ and $\exp A$ has the two elementary divisors $(\lambda \pm 1)^r, (\lambda \pm 1)^r$. In this last case, where r is odd, $\exp A$ is a matrix C with only two equal elementary divisors, both of type (iv). It is important to notice that the two ρ 's associated with the elementary divisors must be equal.¹²

Finally, if A is of type (δ), $\exp A$ is a matrix C of type (iv), with the

¹¹ Cf. I, page 906.

¹² Cf. I, pages 907 and 908.

single elementary divisor $(\lambda - 1)^{2k+1}$. Once again there are two distinct canonical pairs $\approx (H, \exp A)$.

The simple matrices C , which do not have an exponential representation are therefore,

(a) *a matrix C with the single pair of real elementary divisors $(\lambda - a)^r$, $(\lambda - a^{-1})^r$, where a is negative and distinct from minus one;*

(b) *a matrix C with the single elementary divisor $(\lambda + 1)^{2k+1}$;*

(c) *a matrix C with the pair of elementary divisors $(\lambda + 1)^{2k+1}$, where one of the associated ρ 's has the value plus one and the other the value minus one.*

An illustration of (c) is the following.

If $C = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ and $H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, C and H are in canonical form. The elementary divisors of C are $(\lambda + 1)$ and $(\lambda + 1)$, and the first associated ρ has the value $+1$ and the other the value -1 . Let G be the real skew symmetric matrix $G = \begin{pmatrix} 0 & g \\ -g & 0 \end{pmatrix}$. Then $\exp(HG) = \exp \begin{pmatrix} 0 & g \\ g & 0 \end{pmatrix}$ and this cannot be equal to C . On the other hand, if $H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, so that the two associated ρ 's have both the same value plus one, $C = \exp \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & \pi \\ -\pi & 0 \end{pmatrix}$.

In this last simple example a change in the associated ρ 's altered the value of H . That this does not always happen is shown by the following example. Let $C = [1, -1, -1]$ and $H = [-1, -1, 1]$. The elementary divisors of C are $(\lambda - 1)$, $(\lambda + 1)$, $(\lambda + 1)$. The first ρ associated with $\lambda + 1$ has the value -1 and the other the value $+1$. Therefore there is no real skew symmetric matrix G , such that $C = \exp(HG)$. On the other hand, the matrix H has the same elementary divisors as C and $[-1, -1, 1] = \exp(HG)$, where $G = \left[\begin{pmatrix} 0 & \pi \\ -\pi & 0 \end{pmatrix}, 0 \right]$.

Comparison of the canonical forms clearly shows that case (c) is a limiting case of case (a) as a tends to minus one.

Since a matrix C , whose only elementary divisors are two equal pairs $(\lambda - a)^r$, $(\lambda - a^{-1})^r$, always does have a real exponential representation of the form (1), when $a \neq -1$, we have the theorem;

THEOREM 5. *Let C be a proper real automorph of the real non-singular symmetric matrix H . Then the matrix C has a real exponential representation of the form $C = \exp(HG)$, with a skew-symmetric G , if and only if, every real elementary divisor of the form $(\lambda - a)^r$ where a is negative, occurs an*

even number of times among the elementary divisors of C and, when $a = -1$ and r is odd, the number of positive ρ 's associated with $(\lambda - a)^r$ is even.

We next determine those matrices C which can be obtained as limiting cases of matrices with an exponential representation. Let C be of type (i), where a is negative, and $r = 2k$ is even. Then C_1 is of the form

$$\Gamma_r = \begin{pmatrix} A_r & 0 \\ 0 & (A_r^*)^{-1} \end{pmatrix} \quad \text{and} \quad H_1 = \begin{pmatrix} 0 & E_r \\ E_r & 0 \end{pmatrix},$$

where $A_r = aE_r + U_r$. If $B_r = A_r - b^2 V_r^*$, where V_r is obtained from U_r by replacing the units in the even numbered rows by zero, and $D_r = [B_r, (B_r^*)^{-1}]$, then $\lim_{b \rightarrow 0} B_r = A_r$ and $\lim_{b \rightarrow 0} D_r = \Gamma_r$. The matrix D_r is obviously an automorph

of H_1 and, since all the latent roots of D_r are complex, D_r has a representation of the form $\exp(H_1 G_1)$, as long as b is different from 0. If $r = 2k + 1$,

and $F_r = \begin{pmatrix} B_{2k} & \epsilon \\ 0 & a \end{pmatrix}$, where ϵ is the column vector of dimension $2k$ defined by $\epsilon^* = (0, 0, 0, \dots, 1)$, $\lim_{b \rightarrow 0} F_r = A_r$. The latent roots of F_r are all complex,

except for one, which has the value a . If $K_r = [F_r, (F_r^*)^{-1}]$, K_r is an automorph of H_1 , and has elementary divisors which are all complex except for the two simple ones $(\lambda - a)$ and $(\lambda - a^{-1})$. If C has only the s pairs of elementary divisors $(\lambda - a)^{r_i}$, $(\lambda - a^{-1})^{r_i}$, $i = 1, 2, \dots, s$, where $r_i = 2k_i + 1$, then $C_1 = [\Gamma_{r_1}, \Gamma_{r_2}, \dots, \Gamma_{r_s}]$, and $C_1 = \lim_{b \rightarrow 0} K$, where $K = [K_{r_1}, K_{r_2}, \dots, K_{r_s}]$.

The elementary divisors of K are all complex except for s pairs of simple elementary divisors $(\lambda - a)$, $(\lambda - a^{-1})$. If s is even, K has an exponential representation of the form $K = \exp(H_1 G_1)$. If s is odd, K does not have such a representation and K is not the limit of a matrix which does; for otherwise, a matrix C with the single pair of elementary divisors $(\lambda - a)(\lambda - a^{-1})$ would be the limit of a matrix with an exponential representation, and this is impossible. Consequently, if C is a matrix with only elementary divisors of the form $(\lambda - a)^r$, $(\lambda - a^{-1})^r$, where a is negative, C is the limit of matrices of the form $\exp(HG)$, if, and only if, the total number of pairs of elementary divisors for which r is odd, is even.

We now consider matrices C with elementary divisors of the form $(\lambda + 1)^r$. If C is a proper automorph of H , with only s elementary divisors of the form $(\lambda + 1)^{2k_i+1}$, by the same argument that was used in the complex case, (§ 3), $C = \lim D$, where all the elementary divisors of D are complex except for s simple elementary divisors $(\lambda + 1)$. With each elementary divisor of C of the form $(\lambda + 1)^{2k_i+1}$, there is, in the canonical form, associated a $\rho = \pm 1$. Hence with C there is associated a set of $s\rho$'s. In a canonical form for D, H ,

with the s elementary divisors $\lambda + 1$ of D is associated the same set of $s\rho$'s. Since C is a proper automorph of H , s is even and D has an exponential representation, if the number of positive ρ 's in the set is even. If the number of positive ρ 's is odd D does not have such a representation nor is D the limit of a matrix which does. For otherwise, the matrix $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ would be a limit of matrices of the form $\begin{pmatrix} 0 & \exp d \\ \exp d & 0 \end{pmatrix}$. On combining these results we have the theorem:

THEOREM 6. *Let C be a proper real automorph of the real symmetric matrix H . Then C has a real exponential representation of the form (1) or is the limit of real automorphs which do, if, and only if, when a is a negative latent root of C , the total number of elementary divisors of C of the form $(\lambda - a)^{2k+1}$ is even and, when $a = -1$, the total number of positive ρ 's associated with these elementary divisors is even.*

Finally, by the same proof as that of Theorem 4, we have

THEOREM 7. *If C is a real automorph of the real symmetric matrix H and C does not have an exponential representation of the form (1), there exists a real automorph D of H , such that DC does have such a representation. The automorph D is of period two.*

If C is improper, D is improper and D cannot have an exponential representation. But, even when C is proper, it is not possible in every case to find a D , which does have an exponential representation. This is best shown by a simple example. Let

$$C = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

If D is to be proper and of period two, D must be $\pm C$. Therefore $D = C$ and D does not have an exponential representation. On considering the proof of Theorem 4, we see that D has an exponential representation if, and only if, the number of positive ρ 's associated with C_s is even. If H is definite, all ρ 's must have the same value and in this case D always has a representation of the form (1).

5. Lorentzian matrices. We now consider in more detail automorphs of the non-singular symmetric matrix H of order n and index $n - 1$. If C is an automorph of H , the elementary divisors of C must all be linear. At most one pair of real elementary divisors $(\lambda - a)$, $(\lambda - a^{-1})$, where $|a| \neq 1$, can appear among the elementary divisors of C . If no such pair occurs,

C must have one elementary divisor $(\lambda \pm 1)$, with which is associated a $\rho = -1$. The ρ 's associated with all other elementary divisors of C all have the value $+1$. We therefore have

THEOREM 8. *Let H be a real non-singular symmetric matrix of order n and index $n-1$. If C is a real proper automorph of H , $C = \exp(HG)$ with a skew-symmetric G , unless C has a pair of elementary divisors $(\lambda - a)$, $(\lambda - a^{-1})$, where a is negative, and, when $a = -1$, the associated ρ 's have different values.*

In other words C has an exponential representation of the form (1) unless C has a pair of elementary divisors $(\lambda - a)$, $(\lambda - a^{-1})$, where a is negative and $|a| \neq 1$ or is the limit, as a tends to -1 , of an automorph, which has such a pair of elementary divisors. From theorem (6) we deduce the Corollary;

COROLLARY 1. *No automorph, which does not have an exponential representation, is the limit of automorphs, which do.*

If n is even and $(H, C) \sim (H_1, C_1)$, we may take

$$(12) \quad H_1 = \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, E_{n-2} \right] \text{ and } C_1 = [C_2, C_3].$$

The matrix C_3 is a diagonal block matrix with blocks of the form ± 1 or $\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$, where $\alpha^2 + \beta^2 = 1$. The matrix C_2 is a diagonal matrix,

$$C_2 = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}.$$

If $a \neq -1$, each elementary divisor $\lambda + 1$ of C is associated with a ρ , which has the value plus one. Therefore C has an exponential representation of the form (1), unless a is negative. When a is negative, each elementary divisor $\lambda + 1$ of $-C$ is associated with a ρ , which has the value plus one. Accordingly $-C$ does have an exponential representation and we have the theorem:

THEOREM 9. *If H is a real non-singular symmetric matrix of even order n and index $n-1$, and, if C is a proper automorph of H , at least one of the matrices C or $-C$ has an exponential representation of the form (1).*

It is of course obvious that no such theorem is true if n is odd.

In conclusion ¹³ we exhibit the canonical forms H_1 , C_1 and the exponential representations of C_1 , when $n = 4$ and $|C_1| = +1$. If

$$H_1 = \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, 1, 1 \right],$$

C_1 is of a single type;

$$C_1 = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & 1/a & 0 & 0 \\ 0 & 0 & \cos \theta & \sin \theta \\ 0 & 0 & -\sin \theta & \cos \theta \end{pmatrix}.$$

The matrix $C_1 = \exp(H_1 G_1)$, where

$$G_1 = \begin{pmatrix} 0 & -\log a & 0 & 0 \\ \log a & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta \\ 0 & 0 & -\theta & 0 \end{pmatrix}, \text{ when } a \text{ is positive.}$$

As remarked earlier, if a is negative, C_1 does not have such an exponential representation.

THE JOHNS HOPKINS UNIVERSITY.

¹³ Cf. C. C. Macduffee, *The Theory of Matrices* (Berlin, 1933), page 68; F. D. Murnaghan, "On the representation of a Lorentz transformation by means of two-rowed matrices," *American Mathematical Monthly*, vol. 38 (1931), pp. 504-511.

UNBOUNDED CONVEX POINT SETS.*

By J. J. STOKER.

1. Introduction. We are concerned here with the properties of unbounded convex point sets S in three-dimensional Euclidean space E^3 , though many of the theorems and proofs are obviously valid in E^n . In addition to convexity we make the following assumptions on the sets S : a) S is not the entire E^3 , so that boundary points of S exist, b) S is assumed to possess *inner* points in E^3 , c) S is a *closed* set. Frequent use will be made of the following well-known property of all convex sets (bounded or unbounded): there exists on every boundary point of S at least one *support plane* (German: Stützebene) of S , that is, a plane containing the boundary point and having the property that S lies entirely in one of the two closed half-spaces bounded by the plane.¹

A special case of unbounded convex sets, the *convex cone*, is treated in some detail because of its importance in the discussion of the general sets S . By a convex cone we mean a closed convex set C consisting of infinite half-rays all emanating from the same point O , the vertex of the cone. However, in dealing with the cones C it is not convenient to assume that C must possess inner points in E^3 or even in E^2 , but we explicitly omit the case in which C is the entire E^3 . It is hence clear that the vertex O of C is always a boundary point of C relative to E^3 . The terms *inner* and *boundary* point are used, however, in dealing with the cones C , in relation to the dimensionality of C , even though C is always considered as laid in E^3 . This terminology is free of ambiguity since C is convex.²

With every set S there is associated a unique cone C , the *characteristic cone* of S , defined as follows: Through any point $p \in S$ all infinite half-rays $r \subset S$ are drawn. The resulting point set is shown to be not empty and to be closed and convex, i.e. it is a cone C . Two such cones erected on different points $p_i \in S$ are shown to differ only by a translation. The characteristic cone plays a central rôle in the discussion of the sets S .

Most, though not all, of the theorems on convex cones as well as the idea of the characteristic cone are contained in the paper of Steinitz: "Bedingt konvergente Reihen und konvexe Systeme," *Journ. f. d. reine u. angew. Math.*,

* Received March 8, 1939.

¹ Bonnesen-Fenchel, *Theorie der konvexen Körper* (1934), p. 4. The proof given here applies to bounded sets, but could be extended easily to unbounded sets.

² Bonnesen-Fenchel, *loc. cit.*, p. 2.

Bd. 143 (1913); Bd. 144 (1914); Bd. 146 (1915). The theorems of Steinitz are formulated in a terminology which is not suited to our purposes, and since his proofs can be replaced by others quite concise, we shall include proofs of these theorems in order to preserve the continuity of the discussion.

It is well known that the set of boundary points of a *bounded* convex set with inner points in E^3 is homeomorphic with the surface of the sphere. The corresponding problem for unbounded sets S is more complex, as the following simple examples indicate: 1) The space between and on a pair of parallel planes—the set of boundary points is not connected, 2) a solid right circular cylinder with an entire straight line as axis—the set of boundary points forms an open cylinder, 3) a half-space—the set of boundary points is a plane. One of our main purposes will be to show that *these examples exhaust all possibilities in so far as the topological structure of the boundary points of S is concerned.*

The *spherical image* of a convex set is defined in the usual manner: The outward normals on all support planes of S are displaced parallel to themselves and erected at a point O ; the points of intersection of such normals with the unit sphere having O as center constitute a set I , the spherical image of S . In the case of bounded convex sets, I is the entire surface of the sphere. We prove that I of any unbounded set S lies on a closed hemisphere, that every *inner* point of the spherical image I_c of the characteristic cone C of S is a point of I , and give conditions under which I of S is a closed or an open set.

At the close of this paper we prove the following theorem: There exists a support plane T of S on a certain boundary point b of S such that the infinite half-ray taken along the inward normal to T at b lies entirely in S if, and only if, the set of boundary points of S is homeomorphic with the plane. We obtain also a sufficient condition that the set of boundary points of S may possess a representation in the form $z = f(x, y)$ with f one-valued and continuous.

2. Convex cones. To the definition of the convex cone C already given above we add the definition of the cone C_p polar to C : On every support plane of C (all of which evidently contain the vertex O of C) we erect the normal turned away from C at O . The totality of all such normals (considered as infinite half-rays) forms the polar cone C_p of C . Since O is always a boundary point of C (relative to E^3) it is clear that C_p is not an empty set. It is also easy to show that C_p is closed and convex, i. e. it is also a convex cone. (See, for example, Bonnesen-Fenchel, *loc. cit.*, p. 4).

We begin the discussion of the cones C with

THEOREM I. *The polar cone C_{pp} of the polar cone C_p is identical with the original cone ³ C .*

It is clear that the above definition of the polar cone C_p can be given in the following form: the necessary and sufficient condition that a half-ray r_p should belong to C_p is that $\angle(r, r_p) \geq \pi/2$ if r is *any* half-ray belonging to C . If, then, $r \subset C$, it follows at once that $r \subset C_{pp}$ from the above definition of the polar cone; hence we have $C \subset C_{pp}$. On the other hand, if $r \not\subset C$, it is clear that a support plane T of C can be found which will *separate* r from C . (Here we use the fact that C is a closed set.) The outward normal r_p on T at the vertex O of C belongs to C_p by the definition. Hence $\angle(r, r_p) < \pi/2$, and r can not belong to C_{pp} . Hence from $r \not\subset C$, follows $r \not\subset C_{pp}$ and we have thus completed the proof that $C_{pp} \equiv C$.

It is convenient to divide the cones C into the following classes, in which C is: (1) an entire straight line (that is, the entire E^1), (2) an entire plane (that is, the entire E^2), (3) any of the remaining possibilities. It is further convenient to subdivide class (3) into three additional sub-classes, i. e., those in which:

(a) C possesses inner points in E^1 , but not in E^2 . C can be only a single infinite half-ray: if C possessed more than one such half-ray, it would either possess inner points in E^2 , since C is convex, or it would consist of an entire straight line, both of which cases are to be excluded.

(b) C possesses inner points in E^2 but not in E^3 . C can be only the convex portion of the plane between and on two infinite half-rays emanating from the same point, since C is convex and can not be the entire plane.

(c) C possesses inner points in E^3 .

It is clear that these classes and sub-classes are mutually exclusive and exhaust all possibilities in the E^3 .

It is of interest to note the nature of the polar cone C_p in each of the above classes. In cases (1) and (2) this is quite simple: C_p for class (1) is the entire plane, evidently, i. e., C_p is of class (2); C_p for class (2) is an entire straight line, i. e., C_p is of class (1).

LEMMA 1. *The polar of a cone of class (3) is itself of class (3).*

This follows from Theorem I: If the polar C_p of a cone C of class (3) were of class (1), say, then C_{pp} would be of class (2) as we have just seen. But $C \equiv C_{pp}$, which shows that our assumption is absurd. It is clear, also,

³ See Steinitz, *loc. cit.*, vol. 144, p. 10.

that C_p could not be of class (2), by the same argument. Since C_p is not empty, it must be of class (3).

From now on, in this section, we shall consider *cones of class (3) only*. Evidently this class is distinguished from the other two by the following property: The vertex O of C is a boundary point of the cones of class (3), but is an inner point in classes (1) and (2), according to our definition of the terms inner and boundary points when applied to cones C . We prove now a theorem on cones of class (3):

THEOREM II. *There exists a support plane T of C with the following property: the inward normal (that is, the normal turned toward C) on T at O lies in C .*

Since C and C_p are both of class (3), by Lemma 1, the point O is a boundary point of both sets. It is obvious from the definition of C_p that C and C_p have no points in common except O . It follows at once that C and C_p possess a common support plane through ⁴ O . The outward normal r_p (relative to C) on T at O belongs to C_p , the inward normal r on T at O to C_{pp} , hence $r \subset C$ since $C \equiv C_{pp}$.

We can present this theorem in a sharper form, as follows:

THEOREM III. *There exists a support plane T of C with the following property: the inward normal on T at O contains (with the exception of O) only inner points of C .*

We know from the preceding theorem that there exists a support plane T and an inward normal r on it at O , such that $r \subset C$. If r contained an inner point of C , then our theorem would be proved, clearly. If r contained no inner point, it would lie in the boundary of C . There would hence exist a support plane T_1 containing r , which would be perpendicular to T . Through O we take a third plane T_2 perpendicular to both T and T_1 . If T_2 contains an inner point p of C , the ray Op lies in the interior of C and, in addition, a plane T_3 through O normal to Op would clearly be a support plane of C : the ray Op would hence possess the required property. If T_2 contained no inner point of C , then T_2 would necessarily be a support plane of C , and C would lie in the convex portion of space between three planes mutually at right angles. In this case it is clear that every ray in the interior of C would have the required property.

3. The characteristic cone. Since any set S is unbounded, there exists an unbounded sequence of points $p_1, p_2, \dots, p_r, \dots$ in S . Consider any point

⁴ Bonnesen-Fenchel, *loc. cit.*, p. 4. The same remark applies here as in footnote 1.

$p \subset S$. S being closed and convex, all line segments $\overline{pp_v}$ as well as all limit points of such segments belong to S . The half-rays $\vec{pp_v}$ cut the unit sphere with p as center in points q_v which possess a limit point q . The half-ray \vec{pq} is made up entirely of limit points of the segments $\overline{pp_v}$. We have then

LEMMA 2. *Every point $p \subset S$ contains at least one infinite half-ray $r_p \subset S$.*

A half-ray with this property we denote from now on as an *axial* half-ray.

LEMMA 3. *If r is an axial half-ray on point $p \subset S$ and \bar{p} any other point of S , the half-ray \bar{r} on \bar{p} parallel to r is also an axial half-ray of S .*

For if p_1, p_2, \dots, p_v is an unbounded sequence of points of r , the line segments $\overline{pp_v}$ lie in S and all points of \bar{r} are limit points of these segments. Hence $\bar{r} \subset S$.

LEMMA 4. *The set C of all axial half-rays emanating from a point $p \subset S$ is a convex cone, the characteristic cone of S .*

That C is convex is shown as follows: Let p_1 and p_2 be any two points of C . We must show that the segment $\overline{p_1p_2} \subset C$. This is evidently the case if the half-rays $r_1 = \vec{pp_1}$ and $r_2 = \vec{pp_2}$ are identical or opposite in direction, or if either p_1 or p_2 is identical with p . In any other case the plane convex sector defined by r_1 and r_2 belongs to S since S is convex. All half-rays in this sector which emanate from p belong to C , and with them the segment $\overline{p_1p_2}$.

From Lemmas 2 to 4 we conclude: *To every point $p \subset S$ there exists a characteristic cone with p as vertex, and all such cones go into one another by translations.*

4. Topological structure of the boundary points of S . Consider any inner point $p \subset S$ and an infinite half-ray r , going out from p , which contains at least one boundary point of S . On going out from p along r one must come upon a *first* such boundary point, say b , since the set of boundary points on r is closed and also bounded on the side toward p . Any support plane P of S on b cuts the ray r at b as it would otherwise contain p , an inner point of S , and this is manifestly impossible. The line segment \overline{pb} is thus the set of points common to S and r , all of them being inner points except the unique boundary point b . We conclude in

LEMMA 5. *A half-ray drawn from an inner point of S contains a unique boundary point of S , or it lies entirely in S and therefore belongs to the characteristic cone of S .*

Let p be an inner point of S , K the unit sphere with center at p , and C the characteristic cone of S with vertex at p . The points of intersection of C with the surface of K we denote by \bar{C} , the remaining points on the sphere by \bar{B} , i. e. \bar{B} is the complement of C relative to the surface of K . From Lemma 5 we conclude: \bar{B} is the set of points formed by the intersection of the surface of K with infinite half-rays drawn from p to the set B of boundary points of S and the correspondence thus set up is one to one. Since the characteristic cone C for any S is independent of the point $p \subset S$ chosen to define it, the set \bar{C} , and with it \bar{B} , is uniquely defined by S . C being a closed set and not the entire E^3 , it follows that \bar{C} is not the entire surface of K and is closed and that \bar{B} is a non-empty open set.

The sets B and \bar{B} are homeomorphic, that is, the correspondence set up between B and \bar{B} by central projection is not only one to one, as we have seen, but also continuous in both directions. We show the continuity first in the direction $\bar{B} \rightarrow B$. Consider any point $\bar{b} \subset \bar{B}$ and let $b \subset B$ be the point corresponding to \bar{b} . We have to show that to any set $\bar{b}_i \subset \bar{B}$ and having \bar{b} as limit point corresponds a set $b_i \subset B$ with b as limit point. To prove this it is sufficient to show that the set b_i possesses a limit point b^* on the half-ray \vec{pb} , for, assuming the existence of b^* , it is clear that $b^* \subset B$ and Lemma 5 shows that b and b^* are identical. The existence of b^* is readily shown: The boundary point b contains a support plane P of S on one side of which b_i as well as p lie; we can construct a finite cone with vertex on p and base on P which will be bounded and contain if not b_i itself at least an infinite sub-sequence b'_i of b_i , from which, if necessary, a further sub-sequence can clearly be taken which will converge to a limit point b^* on \vec{pb} . The existence of b^* is thus assured and with it the continuity in the direction $\bar{B} \rightarrow B$. The continuity in the direction $B \rightarrow \bar{B}$ can be shown in a similar manner; in fact, this is simpler, since it is evident a priori that every infinite set $\bar{b}_i \subset \bar{B}$ possesses a limit point.

To sum up, we have seen that \bar{B} , the complement of \bar{C} relative to the surface of K is uniquely determined by S , and is homeomorphic with the set of boundary points B of S . The problem of determining the possible topological structure of the boundary points of S is thus resolved into the following problem: Determine the possible topological structures of the open sets on the surface of the unit sphere K which are obtained by removing from the surface of K its intersection with any convex cone with vertex at the center of K .

Before continuing with the solution of this problem, it is of interest to note that Theorem I, Lemmas 2 to 5, and all that we have shown in this section with the proofs, as given, are valid for the E^n , independent of n .

We consider the classification of cones C given in section 2 above. These classes were those in which C is: (1) an entire straight line, (2) a plane, (3) all other cases. In case (1), \bar{C} is made up of two diametrically opposite points of the sphere and \bar{B} is homeomorphic with an open cylinder. In the second case \bar{C} is a great circle on the sphere and \bar{B} is homeomorphic with two distinct planes.

In the third case it is convenient to consider a further division of the cones into sub-classes in which C possesses: (a) inner points in E^1 but not in E^2 , (b) inner points in E^2 but not in E^3 , (c) inner points in E^3 . In section 2 we saw that C in case (3a) is a single half-ray; hence \bar{C} is a single point and its complement \bar{B} is homeomorphic with the plane. We saw also that C in case (3b) is a convex sector of the plane; \bar{C} is a closed segment of a great circle and again \bar{B} is homeomorphic with the plane.

In the case (3c) it is clear that \bar{C} possesses inner points relative to the surface of the unit sphere. We shall show that \bar{C} is convex on the sphere, that is, \bar{C} lies on a hemisphere and with any two of its points also contains a great circle arc of length $\leq \pi$ joining the two points. An immediate consequence of this is that \bar{C} is simply connected, and, since it is also a closed set, its complement \bar{B} relative to the surface of the sphere would be homeomorphic with an open circle, hence also with the plane, which is what we wish to show.

We have, then, to show that \bar{C} in the case under consideration is spherically convex. This follows from the convexity of C . The vertex of C (and center of the sphere) is a boundary point of C , a support plane of C at this point exists, and, since $\bar{C} \subset C$, it follows that \bar{C} lies on a hemisphere. Consider any two points p_1 and p_2 which belong to \bar{C} and which are not at opposite ends of a diameter of the sphere (that such points exist is clear since \bar{C} possesses inner points on the sphere). The entire convex plane sector formed by rays from the center of the sphere to p_1 and p_2 belongs to C ; the intersection of the sector with the sphere belongs to \bar{C} : it is thus clear that the shorter great circle arc joining p_1 and p_2 belongs to \bar{C} . If \bar{C} contained no diametrically opposite points, the spherical convexity would be proved. If \bar{C} should contain a pair of diametrically opposite points p_1 and p_2 , it would also contain a point p_3 different from these. The great circle arc joining p_3 with p_1 and p_2 would, as above, belong to \bar{C} . Hence \bar{C} is spherically convex.

With this we have also determined completely the possible topological structures of the boundary points B of sets S in E^3 . Summing up, we have

THEOREM IV. *The set of boundary points of a set S in E^3 possesses one of the three following topological forms: (1) an open cylinder, (2) two distinct planes, (3) a single plane.*

It is of interest to characterize in more detail the sets S in cases (1) and (2) of the above theorem:

Case (1). We have seen that the characteristic cone in this case is a single straight line L . Consider the set of points P common to S and a plane perpendicular to L . Since such a plane contains no axial half-ray, it follows that P is a bounded set. It is also, of course, closed and convex. By Lemma 3 there exists through every point of P one and only one straight line (parallel to L) which lies in S . S is thus an infinite cylinder.

Case (2). The characteristic cone is a plane. The set S can clearly be generated by moving a plane parallel to itself through a finite distance. S is therefore the space between two parallel planes.

We can now state an evident corollary to Theorem IV: *If the set S in E^3 is not the space between two parallel planes, the set of its boundary points is a connected set.* This corollary is true for sets S in the E^n , and although we confine ourselves here in general to sets in E^3 , it is of interest to give a proof of this fact ⁵ valid in the E^n . This is readily done with the aid of the following:

LEMMA 6. *Consider a closed unbounded convex set S with inner points, all of whose boundary points lie on a pair of parallel planes T_1 and T_2 with at least one boundary point on each plane. Then S is the space between the parallel planes.*

From the Lemmas 2 to 5 (valid for any dimension) we conclude: the characteristic cone of S is a plane parallel to T_1 and T_2 and all points of T_1 and T_2 are boundary points of S . This proves the lemma.

THEOREM V. *Let S be any closed unbounded convex set with inner points which is not the space between parallel planes nor the entire space. Any two boundary points of S can be connected by a continuous plane curve lying in the boundary of S .*

Let b_1 and b_2 be any two boundary points of S , T_1 and T_2 support planes on these points. We may clearly assume without loss of generality that T_1 and T_2 are different. Two cases are to be distinguished: (a) T_1 and T_2 are not parallel, (b) T_1 and T_2 are parallel.

(a). T_1 and T_2 intersect. S lies in the convex portion of space bounded by T_1 and T_2 which contains $\overline{b_1 b_2}$. Consider a two-dimensional plane containing b_1 and b_2 and cutting the intersection of T_1 and T_2 in point a . We are free to assume that $\overline{b_1 b_2}$ contains an inner point $p \subset S$, since otherwise $\overline{b_1 b_2}$

⁵ This result, but not our Theorem IV, is due to Steinitz, *loc. cit.*, vol. 146, p. 10.

itself belongs to the boundary of S . From p lines are drawn to all points of the segments $\overline{b_1a}$ and $\overline{ab_2}$. Every such line clearly contains a single boundary point of S , and we see without difficulty that b_1 and b_2 are connected by a continuous plane curve lying in the boundary of S . (See the proof for homeomorphism of the sets B and \bar{B} in section 4).

(b). T_1 and T_2 are parallel. Lemma 6 and our fundamental assumption insure the existence of a boundary point b which does not lie on T_1 or T_2 and which must lie between T_1 and T_2 , S being convex. Any support plane of S on b must clearly intersect T_1 and T_2 . The proof that b_1 and b_2 are connected by a plane curve lying in the boundary of S may now be conducted in exactly the same way as in (a).

5. The spherical image of S . Consider any boundary point b of S , together with the characteristic cone C and its polar cone C_p erected at this point. Any support plane T of S on b must be a support plane of C also: it follows at once from the definition of C_p given in section 2 that the outward normal n on T must lie in C_p . We erect C and C_p with their common vertex O at the center of the unit sphere and denote by \bar{C}_p the intersection of C_p with the surface of the sphere. From the definition of the spherical image I of S given in section I and from the foregoing we conclude: $I \subset \bar{C}_p$.

We proceed to investigate the relations between I and \bar{C}_p in more detail. Since \bar{C}_p is the spherical image of C , we are in effect investigating the relations between the spherical image of S and that of its characteristic cone. It is convenient to introduce at this point the same classification of convex cones discussed in section 2 and apply it here to the polar cone C_p of the characteristic cone of S :

(1) C_p is an entire straight line, C is a plane, and S the space between two parallel planes. In this case I and \bar{C}_p are evidently identical.

(2) C_p is a plane, C is an entire straight line perpendicular to it, S an infinite cylinder erected on a bounded convex plane set and again $I \equiv \bar{C}_p$ (each is a great circle on the unit sphere). We use here the known result that the "circular image" of a bounded convex set in the plane is the circle: circular image of the plane section of S and spherical image of S are identical.

(3) All other cases, subdivided as follows:

(a) C_p possesses inner points in E^1 but not in E^2 . We know that C_p consists of a single infinite half-ray. The cone C is thus a half-space and one can easily show that S must also be a half-space. It is then evident that $I \equiv \bar{C}_p$.

(b) C_p possesses inner points in E^2 but not in E^3 . C_p is the convex space between two half-rays. The cone C is a wedge: that is, the convex space between and on two half-planes having a common boundary. (The two planes are at right angles to the boundary half-rays of C_p). The set S is easily seen to be an infinite cylinder with generators parallel to the edge of the wedge and having as base an *unbounded* convex set in the plane, in contrast with case (2) of Theorem IV in which the cylinder is erected on a bounded set. The set \bar{C}_p is a closed segment of a great circle (at most a semi-circle) and the spherical image of S is clearly the same as the circular image of the plane, unbounded, convex set upon which the cylinder is erected. It is convenient to defer further consideration of this case until after discussion of the next case, which is exactly analogous for three dimensions.

(c) C_p possesses inner points in E^3 . \bar{C}_p possesses inner points relative to the surface of the sphere. We shall show that *every inner point of \bar{C}_p is a point of the spherical image of S* . Let s be an inner point of \bar{C}_p , P the plane through the common vertex O of C and C_p which is perpendicular to the line joining O with s . P is (1) a support plane of the cone C which (2) contains no point of C except its vertex O ; (1) P is a support plane of C by the definition of C_p , and (2) P contains no point of C except O , since it would otherwise contain an entire half-ray $r \subset C$ which would clearly make an angle $< \pi/2$ with some of the half-rays of C_p which pass through the points of a neighborhood of the inner point $s \in \bar{C}_p$, in contradiction with the definition of C_p . Consider next any inner point $p \in S$ and the characteristic cone C of S erected on p as vertex. The set of points common to S and that one of the two half-spaces bounded by a plane T parallel to P and not containing C we denote by S' . S' is evidently convex; it is moreover *bounded*, since every infinite half-ray emanating from p and lying in the half-space containing S' contains a boundary point of S because of the choice of T . The set S' possesses inner points in E^3 (since p is an inner point of S) and T is a support plane of this set. By a well known theorem on bounded convex sets there exists a second support plane T' of S parallel to T . T' is clearly also a support plane of S . The outward normal on T' is parallel to \vec{Os} (the direction of the outward normal on T relative to C). Hence point s belongs to the spherical image of S , as was to be shown.

We can now consider case (b). The method of proof used for case (c) is not valid here without change, since no support plane of C exists which contains only the vertex of C because of the fact that C contains an entire straight line in its boundary—the “edge” of the wedge—and every support plane of C must contain this line. However, as remarked above, the spherical

image of S is the same as the circular image of the unbounded convex plane set which consists of the intersection of S with a plane perpendicular to the edge of the cone C . We might call this set S^2 . A discussion exactly analogous to that of the above but carried out one dimension deeper would show that every inner point of \bar{C}_p would be a point of the circular image of S^2 . (The term inner point means, of course, inner point relative to the great circle which contains \bar{C}_p). We sum up these results in

THEOREM VI. *The spherical image I of S is contained in the spherical image \bar{C}_p of its characteristic cone. If \bar{C}_p is one-dimensional or two-dimensional, at least every inner point of \bar{C}_p (this term being used with reference to the dimensionality of \bar{C}_p) is a point of I . In all other cases I is identical with \bar{C}_p .*

The cone C_p possesses always a support plane through its vertex. From this we deduce an evident corollary to the above theorem: *I of S lies on a hemisphere.*

At this point a natural question arises: Under what conditions will I be an open set (that is, contain *none* of the boundary points of \bar{C}_p), or a closed set? The cases in which this question remains open are the cases (3b) and (3c) above, I being identical with \bar{C}_p in all the others. Additional restrictive assumptions must be made on the sets S in order to answer this question definitely: Consider the convex set which is bounded by a paraboloid of revolution: the spherical image of this set is evidently an open hemisphere; on the other hand, a semi-infinite right circular cylinder possesses a closed hemisphere as spherical image, though the cone C and with it C_p are the same for both sets—a single half-ray and a half-space. An example of a set for which I is neither open nor closed could easily be given.

We begin our discussion of this problem with the case (3b), which reduces to the consideration of an unbounded convex set S^2 in the plane whose characteristic cone is the convex sector between two half-rays. If C^2 of S^2 is in particular a half-plane, S^2 is also a half-plane, as one readily shows. In this case the circular image I of S^2 and the circular image \bar{C}_p of C^2 are evidently the same—a single point on the unit circle. If C^2 is not a half-plane, which we assume from now on, then \bar{C}_p clearly possesses inner points in E^1 , all of which as we have seen belong to I . We wish to obtain conditions on S^2 which determine whether the two boundary points of \bar{C}_p belong to I or not. That \bar{C}_p has only two boundary points is sufficiently clear. It is of importance to note explicitly that the boundary points of \bar{C}_p are determined by the two boundary rays of C_p and that the latter are the two half-rays at right angles to the boundary rays of C .

Suppose that a boundary point $\bar{b} \in \bar{C}_p$ belongs to I . As remarked above,

the half-ray \vec{Ob} is perpendicular to an axial ray $\vec{r} \subset C$; at the same time \vec{Ob} is parallel to the outward normal on a certain support line L of S^2 . L contains a half-ray (parallel to \vec{r}) by Lemma 3, all points of which must be *boundary* points of S^2 since L is a support line. We conclude: If S contains no infinite half-ray in its boundary, I contains no boundary point of \bar{C}_p , i. e., I is an open set.

On the other hand, suppose that there exists a circle with radius sufficiently large that all boundary points of S^2 outside of the circle lie on infinite half-rays belonging to the boundary of S^2 . We may choose a circle with this property with its center at an *inner* point $p \subset S^2$. On p as vertex we erect the cone C . The boundary rays of C intersect the circle say in points q_1 and q_2 . The angle between \vec{pq}_1 and \vec{pq}_2 is definitely $< \pi$, since C is not a half-plane. The outward normal \vec{n} to C (regarded as a half-ray) erected at q_1 must contain a boundary point b of S^2 (since $\vec{n} \not\subset C$), which must, in addition, lie outside the circle, \vec{n} being a tangent to the circle. By our assumption, there exists a certain half-ray \vec{r}_b starting from b which lies entirely in the boundary of S^2 . It follows that \vec{r}_b lies on a support line of S^2 , and, by Lemma 3, \vec{r}_b must be parallel to a ray in the boundary of C ; if it were in the interior of C all points of \vec{r}_b except b would be inner points of the characteristic cone C erected on b as vertex and hence also inner points of S^2 , which is not possible. Moreover, \vec{r}_b must be parallel to \vec{pq}_1 ; if it were parallel to \vec{pq}_2 (the only other possibility) it would necessarily intersect \vec{pq}_1 , since the angle between \vec{n} and \vec{r}_b would be $> \pi/2$, evidently. Again we see that \vec{r}_b would contain inner points of C and consequently also of S^2 , which is impossible. Hence \vec{n} is perpendicular to \vec{r}_b . It follows at once that the point of I corresponding to \vec{n} is one of the two boundary points of \bar{C}_p . In the same way, by considering an outward normal to C at point q_2 , one shows that the other boundary point of \bar{C}_p also belongs to I . I is therefore closed.

We pass to consideration of case (3c), that in which \bar{C}_p possesses inner points relative to the surface of the unit sphere. Theorem III insures the existence of a support plane T of C_p with the following property: the inward normal \vec{n}_i on T at the vertex O of C_p contains (with the exception of O) only inner points of C_p . Since the polar cone C_{pp} of C_p is identical with C (Theorem II), it follows that the outward normal \vec{n}_o on T at O lies in C . Because of the fact that \vec{n}_i lies in the interior of C_p , the plane T contains no point of C except O . Let p be the point of intersection of the inward normal on T with

the surface of the unit sphere: p is thus an inner point of \bar{C}_p and hence a point of I by Theorem VI. Consider any boundary point of \bar{C}_p , say q . The points p and q determine a great circle on the sphere, a closed and connected segment of which belongs to \bar{C}_p , since C_p is convex. One boundary point of this segment is q while p lies in its interior.

We consider the set S_p consisting of the orthogonal projection of *all* points of S on the plane P determined by the points O , p , and q . (As \bar{C}_p lies on a hemisphere with p as an inner point, it is clear that p and q do not lie on a diameter of the sphere and O , p , and q determine a unique plane). S_p is unbounded, since P contains an axial half-ray of S , i. e. the half-ray in the direction \vec{n}_0 . S_p clearly possesses inner points in the plane, but is not the entire plane, as S possesses a support plane parallel to T (p being a point of I), which is by construction perpendicular to P . As we have seen, the plane T contains no ray of the characteristic cone C of S ; consequently the intersection of S with any plane T_i parallel to T is a closed *bounded* set. Since the T_i are perpendicular to the plane of S_p , we conclude that the correspondence established between S and S_p is such that *to each boundary point of S_p corresponds at least one boundary point of S* . S being unbounded, it is not obvious that S_p , the projection of S on a plane, is a closed set even though S is closed. This we show as follows: Let p be a limit point of a set $p_v \subset S_p$. We must prove that p is the projection of some point of S on the plane P . Consider a circle with p as center which contains an infinite number of points p'_v of the points p_v . As we have seen, the intersection with S on any plane T_i perpendicular to P is a bounded set. It follows that the intersection of S with the right cylinder erected over the circle with p as center is also bounded; a set of points in S corresponding to the $p'_v \subset S_p$ must then possess a limit point p_s on the projection ray through p and p_s belongs to S since S is closed. This establishes the fact that S_p is closed. *The projection S_p of S on the plane P is thus a closed unbounded convex set such that to each boundary point of S_p corresponds at least one boundary point of S . Also, support lines of S_p correspond to support planes of S perpendicular to the plane of S_p and vice versa.*

We may now apply to S_p the reasoning used above for the sets S^2 . The assumption that q , a boundary point of \bar{C}_p , belongs to I of S and consequently also to the circular image I_p of S_p insures the existence of an infinite half-ray in the boundary of S_p , as we have seen, and hence the existence of an unbounded set of boundary points of S all lying in a certain support plane T_1 of S perpendicular to the plane of S_p . The intersection S_1^2 of T_1 with S being convex, it follows that S_1^2 contains an axial half-ray. Since q is any boundary point of I of S we may conclude: *I of S is an open set if S possesses no half-ray in its boundary.*

On the other hand, assume now that a sphere exists such that every support plane of S on any boundary point outside the sphere contains at least one half-ray in common with S . Let b_p be any boundary point of S_p which corresponds to a boundary point b of S outside such a sphere. We show that b_p lies on an infinite half-ray r_1 in the boundary of S_p : There exists a support plane T_1 of S on b perpendicular to the plane of S_p . T_1 contains an infinite half-ray r belonging to the boundary of S by our assumption, and r can not be perpendicular to the plane P containing S_p since, as we have seen, the characteristic cone C of S is such that none of its rays are at right angles to P . It follows that the projection of r on P is an infinite half-ray which evidently lies in the boundary of S_p . The spherical image of S_p is thus a closed set, as we have seen above, and we conclude that the point q in the boundary of \bar{C}_p belongs to I of S . The point q being any boundary point of \bar{C}_p , it follows that all boundary points of \bar{C}_p belong to I and I is closed. We thus have

THEOREM VII. *If I of S is of dimension two, it is (a) an open set if S contains no infinite half-ray in its boundary, (b) a closed set if every support plane of S outside a certain sphere lies on an infinite half-ray belonging to the boundary of S .*

If I of S is of dimension one, the above theorem does not hold without modification, as the following example shows: Consider the set S consisting of the convex portion of the plane bounded by a parabola together with all straight lines through such points at right angles to the plane. The spherical image I of S is evidently an *open* semi-circle, though the condition in (a) of Theorem VII is violated and that of (b) is fulfilled. However, as we have seen in dealing with the sets S^2 above, Theorem VII holds in this case also if it is applied, with obvious changes in the terminology, not to S itself but to the plane section of S which is normal to the plane containing I . We have already seen that I of S in this case is identical with the circular image of such a section.

6. Additional results. We begin this section with a theorem which gives a characteristic property of the sets S whose boundary points are homeomorphic with the plane:

THEOREM VIII. *If the set of boundary points of S is homeomorphic with the plane, there will exist a support plane T on a certain boundary point b of S with the following property: the infinite half-ray taken along the inward normal to T at b (that is, the normal turned toward S) lies entirely in S .*

It is clear that this property is not shared with the sets S whose boundary

points are homeomorphic with a cylinder or a pair of planes, since such sets are *geometrically* right cylinders or the space between parallel planes respectively.

Any set S whose boundary points are homeomorphic with the plane possesses a characteristic cone of class (3), as we have seen in the course of proving Theorem IV. It follows (Lemma I, section 2) that the polar cone C_p of C is also of this class. We may therefore apply Theorem III of section 2 to C_p . This theorem insures the existence of a support plane T_1 of C_p such that the inward normal n on T_1 at the vertex of C_p contains, with the exception of the vertex, only inner points of C_p . The outward normal on T_1 at the vertex lies in C , by Theorem I and the definition of the polar cone. Since n contains an inner point p of \bar{C}_p , it follows that T_1 is parallel to a support plane T of S , since $p \in I$ of S by Theorem VI. The inward normal on T at the boundary point through which it passes is thus parallel to a ray belonging to C and consequently belongs to S , which proves the theorem.

For certain sets S it is possible to find a plane P with the following property: the orthogonal projection of the set B of boundary points of S on P is such that a one to one correspondence between the two sets of points is established. It is clear that only the sets S whose boundary points are homeomorphic with the plane can possess this property. In fact, only *certain* sets of this type possess it, as the example of an infinite half-cylinder shows. We have, however, in the following theorem:

THEOREM IX. *If the characteristic cone C of S possesses inner points in E^3 , a plane P can always be chosen to serve as x, y -plane of a set of orthogonal cartesian coördinates such that the points of the set B will be given by $z = f(x, y)$ with f one-valued and continuous.*

Consider any half-ray r which lies in the interior of C . We show that a plane P at right angles to r has the required property. By Lemma 3, there exists an infinite half-ray parallel to r through every point $b \in B$ which lies entirely in S . This half-ray contains with the exception of b only inner points of S , since r lies in the interior of C . It is moreover clear that b is the only boundary point of S lying on the straight line containing r . This is sufficient to show that the projection of the points of B on P is one to one. One shows also without difficulty that f is continuous: for example, one might use practically the same method as that used at the beginning of section IV. One sees also that the domain of definition of f is the entire plane.

ON THE SMOOTHNESS PROPERTIES OF A FAMILY OF BERNOULLI CONVOLUTIONS.*

By PAUL ERDÖS.

Let $L(u, \sigma)$, $-\infty < u < +\infty$ denote the Fourier-Stieltjes transform, $\int_{-\infty}^{\infty} e^{iux} d\sigma(x)$, of a distribution function $\sigma(x)$, $-\infty < x < +\infty$. Thus if $\beta(x)$ is the distribution function which is 0, $\frac{1}{2}$, 1 according as $x \leq -1$, $-1 < x \leq 1$, $1 < x$, then $L(u, \beta) = \cos u$; and so, if b is a positive constant, $\cos(u/b)$ is the transform of the distribution function $\beta(bx)$. Hence, if a is a positive constant, the infinite convolution

$$\sigma_a(x) = \beta(ax) * \beta(a^2x) * \beta(a^3x) * \cdots$$

is convergent if and only if $a > 1$; its Fourier-Stieltjes transform being

$$(1) \quad L(u, \sigma_a) = \prod_{n=1}^{\infty} \cos(u/a^n), \quad (a > 1).$$

It is known¹ that the distribution function σ_a is continuous for every $a > 1$ and, in fact, is either absolutely continuous or purely singular, depending on the value of a . In this direction it is known² that the set of points x in the neighborhood of which $\sigma_a(x)$ is not constant is either the interval $x \leq a/(a-1)$ or a nowhere dense perfect set of measure zero contained in this interval according as $1 < a \leq 2$ or $2 < a$. While this implies that $\sigma_a(x)$ is singular if $2 < a$ it does not imply that $\sigma_a(x)$ is absolutely continuous if $a < 2$. In fact it has recently³ been shown that there exist certain algebraic irrationalities $a < 2$ for which $L(u, \sigma_a)$ does not tend to zero with $1/u$ and so σ_a cannot be absolutely continuous. (It was conjectured, *loc. cit.*³, that such values of a are clustering at $a = 1 + 0$ which would imply that they lie dense in the interval $1 < a < 2$). On the other hand it is known⁴ that those $a < 2$

* Received July 30, 1939.

¹ B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), 48-88, particularly Theorem 11.

² R. Kershner and A. Wintner, "On symmetric Bernoulli convolutions," *American Journal of Mathematics*, vol. 57 (1935), 541-548.

³ P. Erdős, "On a family of symmetric Bernoulli convolutions," *American Journal of Mathematics*, vol. 61 (1939), 974-976.

⁴ A. Wintner, "On convergent Poisson convolutions," *American Journal of Mathematics*, vol. 57 (1935), 827-838.

for which σ_a is absolutely continuous are certainly clustering at $a = 1 + 0$, since if $a = 2^{1/m}$, where m is a positive integer, then σ_a has a continuous derivative of order $m - 1$.

The object of the present paper is to show that the successive smoothing of σ_a can be considered as the general case when $a \rightarrow 1 + 0$. In fact it will be shown that there exists, for every positive integer m , a positive $\eta(m)$ such that the set of those points a of the interval $1 < a < 1 + \eta(m)$ for which σ_a does not possess a continuous derivative of order $m - 1$ is a set of measure zero. To this end it is sufficient to prove that there exists, for every positive integer m , a positive $\delta(m)$ such that the set of those points a of the interval $1 < a < 1 + \delta(m)$ for which

$$(2) \quad L(u, \sigma_a) = o(|u|^{-m}), \quad u \rightarrow \infty,$$

does not hold is a set of measure zero.

Let c_1, c_2, \dots, c_N be N positive integers which satisfy the following conditions:

- (i) $c_1 \leq 2$;
- (ii) $c_i < c_{i+1}, \quad (i = 1, 2, \dots, N-1)$;
- (iii) $c_{i+1} < 3c_i, \quad (i = 1, 2, \dots, N-1)$;
- (iv) there exists an α such that $2^{\frac{1}{2}} < \alpha < 2$ and $|c_{i+1} - \alpha c_i| < 2, \quad (i = 1, 2, \dots, N-1)$.

LEMMA 1. *There exist two positive absolute constants γ_1, γ_2 such that if M is any fixed number $> \gamma_2$, there are less than $[M^{1/4}]$ different sequences c_1, c_2, \dots, c_N satisfying the requirements (i)-(iv), the inequality $c_N \leq M$, and the condition that the number of those indices i ($i = 1, 2, \dots, N$) which satisfy $|c_{i+1} - \alpha c_i| > \frac{1}{10}$ is less than $\gamma_1 \log M$.*

Proof. Suppose that $|c_{i+1} - \alpha c_i| \leq \frac{1}{10}$ and $|c_{i+2} - \alpha c_{i+1}| \leq \frac{1}{10}$ for a fixed i . Then

$$\left| \frac{c_{i+1}}{c_i} - \alpha \right| < \frac{1}{10c_i},$$

hence

$$\left| \frac{c_{i+1}^2}{c_i} - \alpha c_{i+1} \right| < \frac{c_{i+1}}{10c_i} < \frac{3}{10}$$

by (iii). Consequently, since $|c_{i+2} - \alpha c_{i+1}| < \frac{1}{10}$ by assumption,

$$\left| \frac{c_{i+1}^2}{c_i} - c_{i+2} \right| < \frac{3}{10} + \frac{1}{10} < \frac{1}{2}$$

and so c_{i+2} is uniquely determined as the nearest integer⁵ to c_{i+1}^2/c_i .

⁵ The above considerations are suggested by the investigations of Ch. Pisot, "La répartition modulo un et les nombres algébriques," *Annali d. R. Sc. Norm. Sup. di Pisa*, ser. II, vol. VII, p. 238.

Consequently if i_1, i_2, \dots, i_l denote all those among the N indices i which satisfy the inequality $|c_{i+1} - \alpha c_i| > \frac{1}{10}$ then all indices i which are not of the form $i_r + 1$ or $i_r + 2$ for some $r = 1, 2, \dots, l$, are such that c_i is uniquely determined by c_{i-1} and c_{i-2} . On the other hand, even if j is of the form $i_r + 1$ or $i_r + 2$, so that c_j is not uniquely determined by c_{j-1} and c_{j-2} , then there are, by (iv), (or (i)), at most 4 choices for c_j after c_{j-1} has been determined. Hence there are at most 4^{2l} different sequences c_1, c_2, \dots, c_N which have a given set of exceptional indices i_1, i_2, \dots, i_l .

Finally (ii) and (iv) together with the assumption $a_N \leq M$ clearly imply that $N < 5 \log M$ for sufficiently large M , say for $M > \gamma_2$. Since the number of exceptional indices i_1, i_2, \dots, i_l is less than $\gamma_1 \log M$, by the hypothesis of Lemma 1, it is seen that the number of distinct possible choices for a set of exceptional indices cannot exceed

$$\binom{[5 \log M]}{0} + \binom{[5 \log M]}{1} + \dots + \binom{[5 \log M]}{[\gamma_1 \log M]}$$

and is therefore less than $M^{1/8}$ if γ_1 is chosen sufficiently small. Since it was shown above that there are at most 4^{2l} sequences c_1, c_2, \dots, c_N with a given set of exceptional indices, it follows that the number of distinct sequences c_1, c_2, \dots, c_N which satisfy the requirements of Lemma 1 for a fixed $M > \gamma_2$ is less than

$$M^{1/8} \cdot 4^{2l} < M^{1/8} \cdot 4^{2\gamma_1 \log M} < M^{1/4}$$

if γ_1 is sufficiently small. This completes the proof of Lemma 1.

If a, λ are positive numbers let $A_k = A_k(a, \lambda)$ and $\epsilon_k = \epsilon_k(a, \lambda)$ be defined, for $k = 1, 2, \dots$, by placing

$$(3) \quad \lambda a^k = A_k + \epsilon_k, \quad A_k \text{ integer, } -\frac{1}{2} < \epsilon_k \leq \frac{1}{2}.$$

LEMMA 2. *There exists an absolute constant γ_3 , which shall be chosen to be $> \gamma_2$, such that if M has a fixed value greater than γ_3 , then the measure of the set Γ of those values a in the interval*

$$(4) \quad 2^{\frac{1}{2}} < a < 2$$

for which there exists in the interval

$$(5) \quad 1 < \lambda < 2$$

a $\lambda = \lambda(a)$ such that the inequalities

$$(6.1) \quad \lambda a^k < M; \quad (6.2) \quad |\epsilon_k(a, \lambda)| > \frac{1}{30}$$

hold for at most $\frac{1}{2}\gamma_1 \log M$ distinct values of k , is less than $M^{-\frac{1}{2}}$. It is under-

stood that $\epsilon_k = \epsilon_k(a, \lambda)$ is defined as in (3), and that γ_1, γ_2 are the absolute constants occurring in Lemma 1.

Proof. Suppose, if possible, that Lemma 2 is false. Then there exist at least $[M^{1/4}]$ values of a in (4), say

$$a_j, \quad (j = 1, 2, \dots, [M^{1/4}]),$$

which are in Γ and which are separated by $[M^{1/4}] - 1$ intervals each of which has a length not less than $M^{-3/4}$; so that

$$(7) \quad |a_j - a_k| \geq M^{-3/4}.$$

Since a_j is in Γ , there exists a $\lambda = \lambda(a_j)$ in (5) such that

$$\epsilon_k(a_j, \lambda(a_j)) < \frac{1}{3}0$$

holds for all but $\frac{1}{2}\gamma_1 \log M$ values of k satisfying

$$a_j^k \lambda(a_j) < M,$$

where, according to (3)

$$(8) \quad a_j^k \lambda(a_j) = A_k(a_j, \lambda(a_j)) + \epsilon_k(a_j, \lambda(a_j)) = A_k^{(j)} + \epsilon_k^{(j)}, \text{ say.}$$

It will be shown that

(I) The finite sequence of integers $A_k^{(j)}$ belonging to a fixed j ($= 1, 2, \dots, [M^{1/4}]$) satisfies the hypotheses of Lemma 1 if this sequence of integers is identified with the sequence of integers c_1, c_2, \dots, c_N occurring there; and that

(II) The sequences $A_k^{(j)}$ corresponding to different values of j are distinct. Since there are $[M^{1/4}]$ such sequences this will contradict Lemma 1 and so complete the proof of Lemma 2.

In order to prove (I) notice first that (i), (ii), (iii) are obviously satisfied for $c_i = A_i^{(j)}$. Furthermore, by (8)

$$A_{i+1}^{(j)} + \epsilon_{i+1}^{(j)} = a_j(A_i^{(j)} + \epsilon_i^{(j)})$$

and so, by (3) and (4)

$$|A_{i+1}^{(j)} - a_j A_i^{(j)}| = |a_j \epsilon_i^{(j)} - \epsilon_{i+1}^{(j)}| < 2;$$

so that (iv) is also satisfied, with $\alpha = a_j$. The hypothesis (6.1) assures that the assumption $c_N \leq M$ of Lemma 1 is satisfied. In order to verify the remaining assumption of Lemma 1 recall that there are at most $\frac{1}{2}\gamma_1 \log M$ values of k satisfying (6.1), (6.2). Thus there are at most $\gamma_1 \log M$ values of i such that (6.1), (6.2) are satisfied either for $k = i$ or for $k = i + 1$. But if i has a value distinct from one of these $\gamma_1 \log M$ values, so that

$$|\epsilon_i^{(j)}| < \frac{1}{3}0 \text{ and } \epsilon_{i+1}^{(j)} < \frac{1}{3}0,$$

then, by (4),

$$|A_{i+1}^{(j)} - a_i A_i^{(j)}| = |a_j \epsilon_i^{(j)} - \epsilon_{i+1}^{(j)}| < 1/10.$$

Thus there are at most $\gamma_1 \log M$ indices i for which

$$|A_{i+1}^{(j)} - a_j A_i^{(j)}| > 1/10.$$

This completes the proof of (I).

In order to prove (II), suppose, if possible, that (II) is false. Then there exists a pair of distinct indices j and k such that

$$A_i^{(j)} = A_i^{(k)}$$

for all $i = 1, 2, \dots, N$. Thus, by (3),

$$(9) \quad |a_k^l \lambda(a_k) - a_j^l \lambda(a_j)| < 2$$

holds, for all l such that $a_k^l \lambda(a_k) \leq M$. In particular (9) holds if l is an index for which

$$(10) \quad \frac{1}{4} M > a_k^l > \frac{1}{10} M.$$

Now it may be assumed that $a_k > a_j$ so that, by (7), $a_k \geq a_j + M^{-3/4}$. Then

$$a_k^{l+1} \lambda(a_k) \geq a_k^l \lambda(a_k) (a_j + M^{-3/4})$$

and so, by (9),

$$\begin{aligned} a_k^{l+1} \lambda(a_k) &\geq (a_j^l \lambda(a_j) - 2)(a_j + M^{-3/4}) = a_j^{l+1} \lambda(a_j) \\ &\quad + a_j^l \lambda(a_j) M^{-3/4} - 2(a_j + M^{-3/4}). \end{aligned}$$

Hence, by (5) and (10),

$$a_k^{l+1} \lambda(a_k) \geq a_j^{l+1} \lambda(a_j) + \frac{1}{10} M^{1/4} - 2 - 2(a_j + M^{-3/4}) \geq a_j^{l+1} \lambda(a_j) + 3$$

if M is sufficiently large, say $M > \gamma_3$. Thus

$$|a_k^{l+1} \lambda(a_k) - a_j^{l+1} \lambda(a_j)| \geq 3.$$

This contradicts (9) (since by (10) $a_k^{l+1} \lambda(a_k) < M$) where one could write $l+1$ for l . This contradiction proves (II).

The proof of Lemma 2 is now complete.

LEMMA 3. *There exists, on the interval (4) a zero set Z which has the following property: if a is a point of (4) not contained in Z then there is a positive $\beta = \beta(a)$ such that if M is any fixed number larger than β and if λ is any number in (5), then there are at least $\frac{1}{4}\gamma_1 \log M$ values of k which satisfy both conditions (6.1), (6.2).*

Proof. For any positive integer h let Γ_h denote the set of points a on the interval (4) such that (6.1), (6.2) hold (for some $\lambda = \lambda(a)$ in (5)) for less than $\frac{1}{2}\gamma_1 \log M$ values of k if $M = 2^h$. Then, by Lemma 2,

$$\text{meas } \Gamma_h < 2^{-3h} \text{ if } 2^h > \gamma_3.$$

Thus if Γ_μ denotes for any fixed $\mu > \gamma_3$ the a -set

$$(11) \quad \Gamma \equiv \Gamma_\mu = \sum_{2^h > \mu} \Gamma_h \text{ then } \text{meas } \Gamma_\mu < 4\gamma_\mu^{-1}.$$

It is clear from the definition of Γ , that if a is not in Γ_μ and if $M > \mu$, then, even if M is not of the form 2^h for some h , there are still at least $\frac{1}{4}\gamma_1 \log M$ values of k satisfying (6.1), (6.2) for any value of λ in (5). Thus if a is not in Γ_μ then there is a $\beta = \beta(a)$ satisfying the requirements of Lemma 3; in fact one can choose $\beta = \mu$. Then the set of points a in (4) such that there does not exist a $\beta = \beta(a)$ satisfying the requirements of Lemma 3 is contained in Γ_μ for every positive μ . Thus by (11), Z is a zero set. This completes the proof of Lemma 3.

LEMMA 4. For every $q > 0$ there exists a $\rho = \rho(q) > 1$ and a zero set $Z = Z_q$ of a -values contained in the interval

$$(12) \quad 1 < a < \rho(q)$$

with the following properties: if a is a point of (12) not contained in Z_q then there exists an $\alpha = \alpha(a) > 0$ such that if M is any fixed number greater than α , and if λ is any point of the interval (5), then there are at least $q \log M$ values of k satisfying (6.1), (6.2).

Proof. Let a be a point in the interval $1 < a < 2^{\frac{1}{2}}$ such that no integral power of a is a point of the zero set Z occurring in Lemma 3. Let p_1, p_2, \dots, p_r be those prime numbers such that

$$2^{\frac{1}{2}} < a^{p_1} < a^{p_2} < \dots < a^{p_r} < 2.$$

Now if x is such that $a^x = 2$ then, by the elementary inequalities of Chebyshev, there are two absolute constants γ_4, γ_5 such that

$$(13) \quad \gamma_4 \frac{x}{\log x} > r > \gamma_5 \frac{x}{\log x}.$$

Since a^{p_j} ($j = 1, 2, \dots, r$) is in the interval (4) and not a point of Z , there are, by Lemma 3, for every λ in (5), at least $\frac{1}{4}\gamma_1 \log M$ values of k satisfying

$$(14.1) \quad |\lambda a^{p_i k}| < M, \quad (14.2) \quad |\epsilon_k(a^{p_i})\lambda| > \frac{1}{30}$$

provided $M > \beta(a^{p_i})$. Thus, if $M > \max_{1 \leq i \leq r} \beta(a^{p_i})$, there are at least $\frac{1}{4}\gamma_1 \log M$ values of k satisfying (14.1), (14.2) for each i ($= 1, 2, \dots, r$).

But there are at most $\frac{x \log M}{p_i p_j \log 2}$ values of k such that

$$(a^{p_i p_j})^k = (2^{p_i p_j / x})^k < \frac{1}{\lambda} M < M.$$

Thus there are at least

$$\frac{1}{4} \gamma_1 \log M - \sum_{1 \leq i \leq j \leq r} \frac{\log M}{p_i p_j \log 2}$$

values of k satisfying (6.1) and (6.2). Then by (13) the number of values k which satisfy (6.1) and (6.2) is not less than

$$\frac{1}{4} \gamma_1 \gamma_2 \frac{x}{\log x} \log M - 4 \gamma_2 \frac{x}{(\log x)^2} \log M.$$

But this expression can be made greater than $q \log M$ if x is chosen sufficiently large, i. e., if a is chosen sufficiently small, say $a < \rho(q)$. This completes the proof of Lemma 4 since Z_q may be defined to be the zero set of points a in the interval (12), some integral power of which is a point of Z .

THEOREM. *For every positive integer m , there exists a positive $\delta = \delta(m)$ such that the set of points a of the interval $1 < a < 1 + \delta(n)$ for which*

$$L(u, \sigma_a) = o(|u|^{-m}), \quad u \rightarrow \infty,$$

does not hold is a set of measure zero.

Proof. According to (1)

$$L(u, \sigma_a) = \prod_{n=1}^{\infty} \cos(u/a^n), \quad (a > 1).$$

Thus, if u is in the interval $a^k < u \leq a^{k+1}$

$$L(u, \sigma_a) < \prod_{r=1}^k \cos(a^r(u/a^k)).$$

Now let $\lambda = u/a^k$ so that $1 < \lambda < 2$. Then

$$L(u, \sigma_a) < \prod_{r=1}^k |\cos(\lambda a^r)| = \prod_{\lambda a^r \leq u} |\cos(\lambda a^r)|.$$

By Lemma 4, with $M = u$, if a is chosen in the interval (12) and not in Z_q and if $u > \alpha(a)$ there are at least $q \log u$ factors in this last product which are less than $\cos \pi/30$ so that

$$|L(u, \sigma_a)| < (\cos \pi/30)^{q \log u}, \quad u > \alpha(a).$$

Since, according to Lemma 4, $q (> 0)$ can be chosen arbitrarily this completes the proof of the theorem.

ALGEBRAIC VARIETIES OVER GROUND FIELDS OF CHARACTERISTIC ZERO.*

By OSCAR ZARISKI.

Introduction. In an earlier paper (see footnote ¹⁸) we have derived a number of characteristic properties of *simple points* of an algebraic r -dimensional variety V_r . There the *ground field* K (field of coefficients, or field of constants) was assumed throughout to be *algebraically closed*. In the present paper we generalize our results to any V_r defined by a field Σ of algebraic functions over an arbitrary ground field K of characteristic zero. *We do not assume that K is maximally algebraic in Σ .*

Our generalization has an immediate application to *simple subvarieties* of V_r , of any dimension. This application is given in the last part (V) of the paper. An irreducible s -dimensional subvariety V_s of V_r can be treated as a point \tilde{P} , provided we pass to a new ground field \tilde{K} —a suitable *transcendental* extension of K in Σ —and regard our V_r as an $(r-s)$ -dimensional variety \tilde{V}_{r-s} over \tilde{K} . From our definitions it will follow that V_s is simple for V_r if and only if \tilde{P} is simple for \tilde{V}_{r-s} . The properties of the simple point \tilde{P} yield corresponding properties of the simple V_s . It is this application that should justify (in the eyes of a geometer) our consideration of ground fields which are not algebraically closed.

Let ξ_1, \dots, ξ_n be the coördinates of the general point of V_r and let \mathfrak{o} denote the ring $K[\xi_1, \dots, \xi_n]$. An irreducible V_s on V_r is given by a prime s -dimensional ideal \mathfrak{p} in \mathfrak{o} . Let \mathfrak{S} be the *quotient ring* of V_s ($\mathfrak{S} = \mathfrak{o}_{\mathfrak{p}}$, $a/b \in \mathfrak{S}$ if $a, b \in \mathfrak{o}$, $b \not\equiv 0(\mathfrak{p})$) and let $\mathfrak{P} = \mathfrak{S} \cdot \mathfrak{p}$ be the prime ideal of non units of \mathfrak{S} . We define a *simple* V_s by the condition that *there exist* $r-s$ elements $\eta_1, \dots, \eta_{r-s}$ in \mathfrak{S} such that $\mathfrak{S}(\eta_1, \dots, \eta_{r-s}) = \mathfrak{P}$. The elements η_i are referred to as *uniformizing parameters along V_s , or of V_s* . Our main result concerns the characterization of a simple V_s and of its uniformizing parameters with the aid of the different F'_{ω} of primitive elements ω in \mathfrak{o} . In this characterization we start with an arbitrary set of r elements ξ_1, \dots, ξ_r in \mathfrak{o} such that \mathfrak{o} is integrally dependent on $K[\xi_1, \dots, \xi_r]$. Let F'_{ω} be the different of an element ω in \mathfrak{o} if ξ_1, \dots, ξ_r are taken as the independent variables. Just as a matter of arrangement of the indices it is permissible to assume that ξ_1, \dots, ξ_s are algebraically independent mod \mathfrak{p} . Let $f_i(\xi_1, \dots, \xi_s; \xi_{s+i}) \equiv 0 \pmod{\mathfrak{p}}$ be the irreducible congruence mod \mathfrak{p} which ξ_{s+i} satisfies over $K(\xi_1, \dots, \xi_s)$ ($i=1, 2, \dots, r-s$). We show that *if there exists an element ω in \mathfrak{o} such*

* Received September 28, 1939.

that $F'_{\omega} \not\equiv 0(\mathfrak{p})$, then V_s is simple and the $r-s$ elements $f_i(\zeta_1, \dots, \zeta_s; \zeta_{s+i})$ are uniformizing parameters of V_s ; and conversely.

An almost immediate consequence of this result is that the quotient ring \mathfrak{S} of a simple V_s is integrally closed in Σ .

The burden of the proofs rests naturally on the case of simple points. We consider the residue class field $K_{\mathfrak{p}}$ of a point P , i. e. the field $K_{\mathfrak{p}} = \mathfrak{o}/\mathfrak{p}$. This field is a finite algebraic extension of K . Let K^* be the least normal extension of K which contains $K_{\mathfrak{p}}$. Upon extending the ground field K to K^* , a new variety V^*_r is obtained, and on V^*_r the point P splits into a finite number of points P^*_1, \dots, P^*_h . The most difficult step of the theory is the proof that P is simple for V_r if and only if the points P^*_i are simple for V^*_r , and if the quotient ring \mathfrak{S} of P contains the relative algebraic closure of K in Σ . With the aid of this result, the various theorems concerning the simple point P can be readily deduced from the corresponding theorems concerning the points P^*_i .

This reduction succeeds because at each point P^*_i we have a very special state of affairs, namely the residue class field at each point P^*_i coincides with the new ground field K^* . This is therefore a special case of our problem: it is characterized by the condition $K_{\mathfrak{p}} = K$. This special case is treated first (Part III). Here we pass directly from K to the algebraically closed field determined by K . It is shown that this ground field extension does not cause any splitting of the point P . We then use the results already established in the case of an algebraically closed ground field.

The method just outlined necessitates a preliminary study of the splitting of prime ideals in \mathfrak{o} under algebraic extensions of the ground field (Part I). We could not take over directly the results established in this connection by van der Waerden and Krull, because these authors have only dealt with the special case in which K is maximally algebraic in Σ .

The systematic study of simple points and of simple subvarieties undertaken in this paper is a necessary preliminary to the problem of local uniformization on algebraic varieties which we shall treat in a forthcoming paper.

I. Normal ground field extensions.

1. Let Σ be a field and let K be a subfield of Σ , of characteristic zero. The field K shall be referred to as the ground field. We consider a normal algebraic extension field K^* of K and we wish to show how this extension of the ground field defines a corresponding extension field of Σ , which we shall denote by Σ^* , or by $K^*\Sigma$.

Let Ω be the algebraically closed field determined by K and let K' be the

relative algebraic closure of K in Σ , i.e. the field consisting of all those elements of Σ which are algebraic over K . The fields K' and K^* can be imbedded in Ω . This imbedding is defined to within relative automorphisms of K' and K^* over K , but since K^* is a normal extension of K , the intersection of K' and K^* is a subfield of K' which is independent of the imbedding. Let this subfield be denoted by Δ .

The elements of Σ^* shall be the formal finite sums $\xi^* = a^*_1 \xi_1 + \cdots + a^*_h \xi_h$,¹ $a^*_i \in K^*$, $\xi_i \in \Sigma$, h -arbitrary. Addition, subtraction and multiplication are defined formally in an obvious fashion. We need a rule for identifying two formal sums, and for this it is sufficient to give a rule for identifying a formal sum ξ^* with the zero element of Σ^* . Let $\xi^* = a^*_1 \xi_1 + \cdots + a^*_h \xi_h$ and let b^*_1, \cdots, b^*_n ($b^*_i \in K^*$) be an independent Δ -basis of the algebraic extension field $\Delta(a^*_1, \cdots, a^*_h)$ of Δ . If we substitute formally into the sum $a^*_1 \xi_1 + \cdots + a^*_h \xi_h$ the expressions of a^*_1, \cdots, a^*_h in terms of b^*_1, \cdots, b^*_n (linear forms with coefficients in Δ), we get an expression of the form $b^*_{1\eta_1} + \cdots + b^*_{n\eta_n}$, $\eta_i \in \Sigma$. To indicate this substitution we write: $\xi^* \rightarrow b^*_{1\eta_1} + \cdots + b^*_{n\eta_n}$. We identify the element ξ^* with the zero element of Σ^* , if and only if $\eta_1 = 0, \cdots, \eta_n = 0$. It is self-evident that this identification rule is independent of the choice of the base b^*_1, \cdots, b^*_n . More generally, let c^*_1, \cdots, c^*_m be a set of elements of K^* which are such that: (1) they are linearly independent over Δ ; (2) the a^*_i can be expressed as linear forms of the c^*_j with coefficients in Δ . The elements c^*_j need not belong to the field $\Delta(a^*_1, \cdots, a^*_h)$. By condition (2) we get, through formal substitution: $\xi^* \rightarrow c^*_{1\zeta_1} + \cdots + c^*_{m\zeta_m}$. We assert that $\xi^* = 0$ if and only if $\zeta_1 = \cdots = \zeta_m = 0$. For the proof, let d^*_1, \cdots, d^*_v be an independent Δ -basis of $\Delta(b^*_1, \cdots, b^*_n, c^*_1, \cdots, c^*_m)$ and let $\xi^* \rightarrow d^*_{1\omega_1} + \cdots + d^*_{v\omega_v}$. It is clear that $b^*_{1\eta_1} + \cdots + b^*_{n\eta_n} \rightarrow d^*_{1\omega_1} + \cdots + d^*_{v\omega_v}$ and also $c^*_{1\zeta_1} + \cdots + c^*_{m\zeta_m} \rightarrow d^*_{1\omega_1} + \cdots + d^*_{v\omega_v}$. If $b^*_i = \sum_{j=1}^v k_{ij} d^*_j$, $k_{ij} \in \Delta$, then the matrix (k_{ij}) is of rank n , since b^*_1, \cdots, b^*_n are linearly independent over Δ , and moreover $\omega_i = \sum_{j=1}^n k_{ji} \eta_j$. Similarly, if $c^*_i = \sum_{j=1}^v l_{ij} d^*_j$, then the matrix (l_{ij}) is of rank m , and we have $\omega_i = \sum_{j=1}^m l_{ji} \zeta_j$. Hence, if $\xi^* = 0$, i.e. if $\eta_1 = \cdots = \eta_n = 0$, then $\omega_1 = \cdots = \omega_v = 0$, and since (l_{ij}) is of rank m , it follows that $\zeta_1 = \cdots = \zeta_m = 0$. Conversely, if $\zeta_1 = \cdots = \zeta_m = 0$, then $\omega_1 = \cdots = \omega_v = 0$, i.e. $\sum_{j=1}^n k_{ji} \eta_j = 0$, $i = 1, 2, \cdots, v$, and since the matrix (k_{ij}) is of rank n , it follows that $\eta_1 = \cdots = \eta_n = 0$, i.e. $\xi^* = 0$.

¹We use small Greek letters for elements of Σ and small Latin letters for elements of K . The same letters with an asterisk denote elements of Σ^* and K^* respectively.

As an immediate consequence we have the following: if a^*_1, \dots, a^*_h are themselves linearly independent over Δ , then $a^*_1\xi_1 + \dots + a^*_h\xi_h = 0$ if and only if $\xi_1 = \dots = \xi_h = 0$.

It is clear that formal addition and multiplication of the elements of Σ^* , considered as formal sums, is consistent with out identification rule. Hence Σ^* is a ring.

LEMMA 1. Let θ be an element of K^* and let $f(\theta) = \theta^g + a_1\theta^{g-1} + \dots + a_g = 0$, $a_i \in \Delta$, be the irreducible equation for θ over Δ . The polynomial $f(x)$ remains irreducible in the polynomial ring $\Sigma[x]$; in other words: the relative degrees $[\Delta(\theta) : \Delta]$, $[\Sigma(\theta) : \Sigma]$ are the same.

Proof. Let $\phi(x) = x^m + \omega_1x^{m-1} + \dots + \omega_m$, $\omega_i \in \Sigma$, be an irreducible factor of $f(x)$ in $\Sigma[x]$. Let $\theta^{(1)} = \theta$, $\theta^{(2)}, \dots, \theta^{(m)}$ be the roots of $\phi(x)$. Since $\theta \in K^*$ and since K^* is a normal extension of Δ , all the roots of $f(x)$ are in K^* . Consequently, $\omega_1, \dots, \omega_m \in K^*$. Since the ω 's are in Σ and are algebraic over K , they must also belong to the field K' . Consequently $\omega_1, \dots, \omega_m \in \Delta$, whence $\phi(x) = f(x)$, q. e. d.

By means of this Lemma we now show that Σ^* is an integral domain. (i. e. Σ^* has no zero divisors). Let $\xi^*\eta^* = 0$, $\xi^* = a^*_1\xi_1 + \dots + a^*_m\xi_m$, $\eta^* = b^*_1\eta_1 + \dots + b^*_n\eta_n$, and let g be the relative degree of the field $\Delta(a^*_1, \dots, a^*_m, b^*_1, \dots, b^*_n)$ with respect to Δ . Let θ be a primitive element of this field, satisfying an irreducible equation $F(\theta) = 0$ of degree g , with coefficients in Δ . By our identification rule we have:

$$\begin{aligned} (1) \quad \xi^* &= \alpha_0 + \alpha_1\theta + \dots + \alpha_{g-1}\theta^{g-1} = \phi(\theta), \\ (1') \quad \eta^* &= \beta_0 + \beta_1\theta + \dots + \beta_{g-1}\theta^{g-1} = \psi(\theta), \end{aligned} \quad \alpha_i, \beta_j \in \Sigma,$$

and by the same rule, the relation $\phi(\theta) \cdot \psi(\theta) = 0$ implies that the polynomial $\phi(x) \cdot \psi(x)$ is divisible (in $\Sigma[x]$) by $F(x)$. By Lemma 1, $F(x)$ is irreducible in $\Sigma[x]$. Hence, either $\phi(x)$ or $\psi(x)$ is identically zero, i. e. either $\xi^* = 0$ or $\eta^* = 0$, which shows that Σ^* has no zero divisors.

It now follows immediately that Σ^* is a field. In fact, every element ξ^* of Σ^* is of the form (1), for some $\theta \in K^*$, and, by Lemma 1, Σ^* contains the entire field $\Sigma(\theta)$.

Remark 1. We call attention to the important role which the field Δ plays in the definition of the field Σ^* . It is this field, rather than the ground field K , which really matters in our construction. By definition, Δ is the largest subfield of Σ which can be imbedded in K^* . We would get the same field Σ^* if we took Δ as ground field instead of K .

Of particular importance is the special case $K = K'$ (i. e. K is "maximally algebraic" in Σ , or K is algebraically closed in Σ). In this case we have $K = \Delta$ for every normal extension of K .

Remark 2. The fields Σ and K^* are subfields of Σ^{*2} and have at least the field K in common. It is not difficult to see that Σ^* is the *smallest field having this property*, i. e. any field Γ with this property contains Σ^* . Our hypothesis is to the effect that the field Γ contains two subfields Σ_1 and K^*_{*1} simply isomorphic to Σ and K^* respectively, and, moreover, that the field K_1 which corresponds to K in the isomorphism between K^*_{*1} and K^* is a subfield of Σ_1 . It is then clear that the intersection of Σ_1 and K^*_{*1} must be the field Δ_1 which corresponds to Δ in the isomorphism $\Sigma \cong \Sigma_1$. Using the reasoning of the proof of Lemma 1, it is immediately seen that the join (Σ_1, K^*_{*1}) of the two subfields Σ_1 and K^*_{*1} of Γ is abstractly isomorphic to the field Σ^* , and that this isomorphism induces the given isomorphisms between Σ_1 and Σ , and between K^*_{*1} and K^* .

2. Let \mathfrak{o} be an arbitrary subring of Σ , subject to the only condition: $K \subset \mathfrak{o}$. Let $\mathfrak{o}^* = K^*\mathfrak{o}$ be the extended ring in Σ^* , i. e. the ring whose elements are of the form $a^*_1\xi_1 + \dots + a^*_n\xi_n$, $a^*_i \in K^*$, $\xi_i \in \mathfrak{o}$. Let Δ' be the intersection of \mathfrak{o} with Δ . Since Δ is an algebraic extension of K and since $K \subset \mathfrak{o}$, it follows that Δ' is a field.

THEOREM 1. *If $\Delta' = \Delta$, then $\mathfrak{o}^*\mathfrak{A} \cap \mathfrak{o} = \mathfrak{A}$ for any \mathfrak{o} -ideal \mathfrak{A} . In the general case the relation $\mathfrak{o}^*\mathfrak{A} \cap \mathfrak{o} = \mathfrak{A}$ still holds true if \mathfrak{A} is prime.*

Proof. Let $\xi = a^*_1\xi_1 + \dots + a^*_n\xi_n$, $a^*_i \in K^*$, $\xi_i \in \mathfrak{A}$, be an element of $\mathfrak{o}^*\mathfrak{A} \cap \mathfrak{o}$, and let θ be a primitive element of $\Delta'(a^*_1, \dots, a^*_n)/\Delta'$. Since $\Delta' \subset \mathfrak{o}$, we can write ξ in the form:

$$(2) \quad \xi = \eta_0 + \eta_1\theta + \dots + \eta_{g-1}\theta^{g-1},$$

where $\eta_i \in \mathfrak{A}$ and where g is the relative degree of $\Delta(a^*_1, \dots, a^*_n)$ with respect to Δ' . Under the hypothesis that $\Delta' = \Delta$, the elements $1, \theta, \dots, \theta^{g-1}$ are linearly independent over Δ , and hence the equation (2) implies that $\xi = \eta_0$, $\eta_1 = \dots = \eta_{g-1} = 0$. Hence $\xi \equiv 0(\mathfrak{A})$. This shows that $\mathfrak{o}^*\mathfrak{A} \cap \mathfrak{o} \subseteq \mathfrak{A}$, and since $\mathfrak{A} \subseteq \mathfrak{o}^*\mathfrak{A}$, it follows that $\mathfrak{o}^*\mathfrak{A} \cap \mathfrak{o} = \mathfrak{A}$.

In the general case and for a *prime ideal* \mathfrak{A} ,³ we proceed as follows. Multiplying (2) by $1, \theta, \dots, \theta^{g-1}$ respectively, we get relations of the form:

² More precisely: Σ^* contains two subfields abstractly isomorphic to Σ and K^* respectively, consisting of the elements $a^*_1\xi_1 + \dots + a^*_m\xi_m$ in which $a^*_1, \dots, a^*_m \in K$ or $\xi_1, \dots, \xi_m \in K$ respectively.

³ The following example illustrates the possibility: $\mathfrak{o}^*\mathfrak{A} \cap \mathfrak{o} \neq \mathfrak{A}$, if $\Delta \neq \Delta'$. Let K be the field of rational numbers, $K^* = K(\sqrt{2})$ and let $\Sigma = K^*(x)$. If we regard K as the ground field then the extension $K \rightarrow K^*$ does not affect Σ , i. e. we have $\Sigma = \Sigma^*$. Let $\mathfrak{o} = K[x, x \cdot \sqrt{2}]$, $\mathfrak{A} = \mathfrak{o} \cdot x$. Then $\mathfrak{o}^* = K^*[x]$, $\mathfrak{o}^*\mathfrak{A} = \mathfrak{o}^* \cdot x$ and $\mathfrak{o}^*\mathfrak{A} \cap \mathfrak{o} = \mathfrak{o} \cdot (x, x \cdot \sqrt{2}) \neq \mathfrak{A}$. Here the fields Δ and Δ' coincide with K^* and K respectively.

$$\begin{aligned}\xi &= \eta_0 + \eta_1 \theta + \cdots + \eta_{g-1} \theta^{g-1}, \\ \xi \theta &= \eta_0^{(1)} + \eta_1^{(1)} \theta + \cdots + \eta_{g-1}^{(1)} \theta^{g-1}, \\ &\vdots \\ \xi \theta^{g-1} &= \eta_0^{(g-1)} + \eta_1^{(g-1)} \theta + \cdots + \eta_{g-1}^{(g-1)} \theta^{g-1},\end{aligned}$$

where all the $\eta_j^{(i)}$ are in \mathfrak{A} . Hence $|\eta_i^{(j)} - \delta_i^{(j)} \xi| = 0$, where $\delta_i^{(j)} = 0$ if $i \neq j$, and $\delta_i^{(i)} = 1$. This equation is of the following form:

$$\xi^g + \beta_1 \xi^{g-1} + \cdots + \beta_g = 0,$$

where $\beta_i \equiv 0(\mathfrak{A})$, $i = 1, 2, \dots, g$. Hence $\xi^g \equiv 0(\mathfrak{A})$, and since \mathfrak{A} is prime, we conclude, as in the first part of the proof, that $\xi \equiv 0(\mathfrak{A})$, q. e. d.

An ideal \mathfrak{A}^* in \mathfrak{o}^* is said to lie over an ideal \mathfrak{A} in \mathfrak{o} if the relation $\mathfrak{A}^* \cap \mathfrak{o} = \mathfrak{A}$ is satisfied. It has been proved by Krull⁴ that, over every prime ideal \mathfrak{p} in \mathfrak{o} there lies at least one prime ideal \mathfrak{p}^* in \mathfrak{o}^* , provided that \mathfrak{o}^* be integrally dependent on \mathfrak{o} (i. e. that each element of \mathfrak{o}^* be integrally dependent on elements of \mathfrak{o}). This provision is satisfied in our case, since $\mathfrak{o}^* = K^* \mathfrak{o}$ and since K^* , as an algebraic extension field of K , is certainly integrally dependent on K ($K \subset \mathfrak{o}$).

We consider a prime \mathfrak{o}^* -ideal \mathfrak{p}^* which lies over \mathfrak{p} and we denote by $K_{\mathfrak{p}}$ and $K^*_{\mathfrak{p}^*}$ the residue class fields of \mathfrak{p} and \mathfrak{p}^* respectively, i. e. the quotient fields of the residue class rings $\mathfrak{o}/\mathfrak{p}$ and $\mathfrak{o}^*/\mathfrak{p}^*$ respectively. Since $\mathfrak{p}^* \cap \mathfrak{o} = \mathfrak{p}$, $K_{\mathfrak{p}}$ may be regarded as a subfield of $K^*_{\mathfrak{p}^*}$. Moreover, K and K^* may be regarded as subfields of $K_{\mathfrak{p}}$ and $K^*_{\mathfrak{p}^*}$ respectively.

LEMMA 2. $K^*_{\mathfrak{p}^*}$ is the extension field of $K_{\mathfrak{p}}$ obtained by the extension $K \rightarrow K^*$ of the ground field K ; in symbols: $K^*_{\mathfrak{p}^*} = K^* \cdot K_{\mathfrak{p}}$.

We observe that K^* and $K_{\mathfrak{p}}$ are subfields of $K^*_{\mathfrak{p}^*}$ having at least the field K in common. Hence, by Remark 2 of the preceding section, we have: $K^*_{\mathfrak{p}^*} \supseteq K^* \cdot K_{\mathfrak{p}}$. On the other hand, any element of $\mathfrak{o}^*/\mathfrak{p}^*$ is of the form $a^*_1 \bar{\xi}_1 + \cdots + a^*_m \bar{\xi}_m$, $a^*_i \in K^*$, $\bar{\xi}_i \in \mathfrak{o}/\mathfrak{p}$. This shows that the ring $\mathfrak{o}^*/\mathfrak{p}^*$, and hence also its quotient field $K^*_{\mathfrak{p}^*}$, is contained in the field $(K^*, K_{\mathfrak{p}})$. Hence $K^*_{\mathfrak{p}^*} = K^* K_{\mathfrak{p}}$, as was asserted.

3. Unramified character of the maximal \mathfrak{o} -ideals. We make the following assumption:

The field Δ is a finite extension of K . This assumption is always satisfied if, for instance, K' (the algebraic closure of K in Σ) is itself a finite extension of K .

Under this assumption we prove the following fundamental theorem:

⁴W. Krull, "Zum Dimensionbegriff der Idealtheorie" (Beiträge zur Arithmetik kommutativer Integritätsbereiche, III), *Mathematische Zeitschrift*, vol. 42 (1937), p. 749.

THEOREM 2. If \mathfrak{p} is a maximal \mathfrak{o} -ideal⁵ then $\mathfrak{o}^*\mathfrak{p}$ is the intersection of the prime \mathfrak{o}^* -ideals which lie over \mathfrak{p} .

For polynomial rings $\mathfrak{o} = K[x_1, \dots, x_n]$ this theorem is due to van der Waerden.⁶ It appears as a special case of a generalized discriminant theorem proved by Krull for any pair of integral domains $\mathfrak{o}, \mathfrak{o}^*$ (\mathfrak{o}^* -integrally dependent on \mathfrak{o}) under the hypothesis that \mathfrak{o} is integrally closed in its quotient field.⁷ If we assume, as it is permissible to do, that Σ is the quotient field of \mathfrak{o} , then Krull's hypothesis in our special case implies that $K' \subset \mathfrak{o}$, whence $\Delta = \Delta'$. The special case when \mathfrak{o}^* is obtained from \mathfrak{o} by a separable extension of the ground field has been treated separately by Krull in his report "Idealtheorie" (p. 40). However, also this treatment is based on the tacit assumption that the fields Δ and Δ' coincide. Namely, under this assumption it is permissible to take Δ as ground field, since $\Delta = \Delta' \subset \mathfrak{o}$, i.e. we may put $\Delta = K$, and then our assumption $\Delta = \Delta'$ becomes: $K' \cap K^* = K$. It follows then, by Lemma 1, that the Galois groups of Σ^*/Σ and of K^*/K coincide (i.e. every relative automorphism of K^* over K can be extended to a relative automorphism Σ^* over Σ ; note that Σ^* is at any rate a normal extension of Σ). One defines then in a natural fashion the concepts of *conjugate ideals* and of *invariant ideals* in \mathfrak{o}^* . The proof by van der Waerden and its generalization by Krull are then applicable, leading to the following theorem:

THEOREM 2' (van der Waerden-Krull). If $\Delta = \Delta'$, and if K^* is a separable extension of Δ , then each invariant \mathfrak{o}^* -ideal \mathfrak{A}^* is the extended ideal of its contracted ideal in \mathfrak{o} : $\mathfrak{A}^* = \mathfrak{o}^* \cdot (\mathfrak{A}^* \cap \mathfrak{o})$, and for each prime \mathfrak{o} -ideal \mathfrak{p} ⁸ it is true that $\mathfrak{o}^*\mathfrak{p}$ is the intersection of the prime \mathfrak{o}^* -ideals which lie over \mathfrak{p} .

We shall make use of Theorem 2' in order to prove our more general theorem for maximal ideals.

Let $\bar{\Delta}$ be the least normal extension of K which contains the field Δ , i.e. $\bar{\Delta}$ is the join of Δ and of its conjugate fields over K . By our assumption, $\bar{\Delta}$ is a finite extension of K . We introduce the intermediate field $\bar{\Sigma} = \Delta\bar{\Sigma}$, a finite algebraic extension of Σ , and the intermediate ring $\bar{\mathfrak{o}} = \bar{\Delta}\mathfrak{o}$, so that $\Sigma \subseteq \bar{\Sigma} \subseteq \Sigma^*$, $\mathfrak{o} \subseteq \bar{\mathfrak{o}} \subseteq \mathfrak{o}^*$. The ground field extension $K \rightarrow K^*$ is thus decomposed into two successive normal extensions: $K \rightarrow \bar{\Delta}$, $\bar{\Delta} \rightarrow K^*$. We have clearly the relations: $\Sigma^* = K^*\bar{\Sigma}$, $\mathfrak{o}^* = K^*\bar{\mathfrak{o}}$.

⁵ An ideal is maximal (or divisorless) if it is not properly contained in any other ideal, different from the unit ideal.

⁶ B. L. van der Waerden, "Eine Verallgemeinerung des Bezoutschen Theorems," § 5, *Mathematische Annalen*, vol. 99 (1928).

⁷ W. Krull, "Der allgemeine Discriminantsatz. Unverzweigte Ringerweiterungen" (Beiträge zur Arithmetik kommutativer Integritätsbereiche, VI), *Mathematische Zeitschrift*, vol. 45 (1939).

⁸ Not necessarily maximal as in Theorem 2.

We assert that *the relative algebraic closure $\bar{\Delta}'$ of $\bar{\Delta}$ in $\bar{\Sigma}$ is the field $\bar{\Delta}K'$* . To show this, we first observe that $\Delta = \bar{\Delta} \circ K'$, and hence, by Lemma 1, $[\bar{\Delta} : \Delta] = [\bar{\Sigma} : \Sigma] = g$. Let $\tilde{\alpha}$ be an element of $\bar{\Sigma}$ which is algebraic over $\bar{\Delta}$, hence also algebraic over K' . The relative degree $[K'(\tilde{\alpha}) : K']$ cannot be greater than the relative degree $[\Sigma(\tilde{\alpha}) : \Sigma]$, and since $[\Sigma(\tilde{\alpha}) : \Sigma] \leq [\bar{\Sigma} : \Sigma] = g$, it follows that $[K'(\tilde{\alpha}) : K'] \leq g$. This last inequality holds true for any element α in $\bar{\Delta}'$, and consequently $[\bar{\Delta}' : K'] \leq g$. On the other hand, $\bar{\Delta}'$ contains the field $\bar{\Delta}K'$, and we have $[\bar{\Delta}K' : K'] = [\bar{\Delta} : \Delta] = g$, in view of the relation $\Delta = \bar{\Delta} \circ K'$ and of Lemma 1. Hence necessarily $\bar{\Delta}' = \bar{\Delta}K'$, as was asserted.

We now prove the relation:

$$(3) \quad \bar{\Delta} = K^* \circ \bar{\Delta}K'.$$

Let a^* be an element of $K^* \circ \bar{\Delta}K'$. Since $\Delta = K^* \circ K'$, we have $[K'(a^*) : K'] = [\Delta(a^*) : \Delta]$. Now we have just proved that $[\bar{\Delta}K' : K'] = g$. Since $a^* \in \bar{\Delta}K'$, we conclude that $[\Delta(a^*) : \Delta] \leq g$, for any element a^* in $K^* \circ \bar{\Delta}K'$. Hence this last field is of relative degree $\leq g$ over Δ . Since on the other hand this field contains $\bar{\Delta}$, and since $[\bar{\Delta} : \Delta] = g$, the relation (3) is established.

The relation (3) says that $\bar{\Delta}$ is the intersection of K^* with the algebraic closure of $\bar{\Delta}$ in $\bar{\Sigma}$. The ground field extension $\bar{\Delta} \rightarrow K^*$ therefore satisfies the condition of Theorem 2'. We therefore know that every prime ideal \bar{p} in \bar{o} is the intersection of the prime o^* -ideals which lie over p . Let us assume that Theorem 2 has already been proved for the ground field extension $K \rightarrow \bar{\Delta}$ and, moreover, let us assume that there is only a finite number of prime \bar{o} -ideals which lie over a given maximal prime o -ideal p . We will have then:

$$\bar{o}p = [\bar{p}_1, \bar{p}_2, \dots, \bar{p}_m].$$

The ideals \bar{p}_i are also maximal in o ,⁹ hence are two by two free from common divisors. Therefore their intersection coincides with their product:

$$(4) \quad \bar{o}p = \bar{p}_1 \bar{p}_2 \cdots \bar{p}_m.$$

By Theorem 2' we have

$$o^* \bar{p}_i = [p^*_{i1}, p^*_{i2}, \dots],$$

where $p^*_{ij} \circ \bar{o} = \bar{p}_i$. Since $(\bar{p}_i, \bar{p}_j) = \bar{o}$, if $i \neq j$, we have also $(o^* \bar{p}_i, o^* \bar{p}_j) = o^*$. Hence the product of the ideals $o^* \bar{p}_i$ coincides with their intersection, and therefore, by (4),

⁹ Since p is maximal, the ring o/p is a field. The integral domain \bar{o}/\bar{p}_i is integrally dependent on its subfield o/p and hence is also a field. Consequently \bar{p}_i is maximal.

$$\begin{aligned} o^*p &= o^*\bar{p}_1 \cdot o^*\bar{p}_2 \cdots o^*\bar{p}_m = [o^*\bar{p}_1, \cdots, o^*\bar{p}_m] \\ &= [p_{11}^*, p_{12}^*, \cdots, p_{m1}^*, \cdots], \end{aligned}$$

which proves Theorem 2.

Thus, to complete the proof of Theorem 2, we have only to prove it for arbitrary *finite* ground field extension $K \rightarrow \bar{\Delta}$, and we have also to show that the number of prime \bar{o} -ideals which lie over p is finite. This we shall do in the following section.

4. Merely as a matter of notations, we may identify Δ' with K , since $\Delta' \subset o$. Let p be a maximal o -ideal and let $K_p (= o/p)$ be the residue class field of p . Let $K_p \cap \bar{\Delta} = \Delta_p$,¹⁰ whence $K \subseteq \Delta_p \subseteq \bar{\Delta}$, and let $\Delta_p = K(\vartheta)$, where ϑ is a primitive element of Δ_p over K . Let $f(\vartheta) = 0$ be the irreducible equation, say of degree m , which ϑ satisfies over K . Since $\vartheta \in K_p$ and $K_p = o/p$ (by hypothesis: p is maximal!), there must exist in o an element ω such that

$$(5) \quad f(\omega) \equiv 0(p).$$

If \bar{p} is a prime \bar{o} -ideal which lies over \bar{p} , then

$$(5') \quad f(\omega) \equiv 0(\bar{p}).$$

Since $\bar{\Delta}$ is normal over K and since one root, $\vartheta = \vartheta_1$, of the polynomial $f(x)$ is in $\bar{\Delta}$, all its roots are in $\bar{\Delta}$, whence also in \bar{o} . Hence, by (5'), we must have $\omega \equiv \vartheta_i(\bar{p})$, where ϑ_i is one of the roots $\vartheta_1, \cdots, \vartheta_m$ of $f(x)$. Let, say $\omega \equiv \vartheta_1(p)$. We assert that

$$(6) \quad \bar{p} = (\bar{o}p, \omega - \vartheta_1).$$

Let θ be a primitive element of $\bar{\Delta}$ over $K(\vartheta_1)$, and let $[\bar{\Delta}:K(\vartheta_1)] = n$. Every element $\bar{\alpha}$ of \bar{o} can be written in the form:

$$\bar{\alpha} = \bar{\alpha}_0 + \bar{\alpha}_1\theta + \cdots + \bar{\alpha}_{n-1}\theta^{n-1},$$

where

$$\bar{\alpha}_i = \alpha_{i0} + \alpha_{i1}\vartheta_1 + \cdots + \alpha_{i,m-1}\vartheta_1^{m-1}, \quad \alpha_{ij} \in o.$$

Since $\omega \equiv \vartheta_1(\bar{p})$, we have: $\alpha_i \equiv \alpha_{i0} + \alpha_{i1}\omega + \cdots + \alpha_{i,m-1}\omega^{m-1}(\bar{p})$. The right-hand side of this congruence is an element of o . Consequently, in the homomorphism $\bar{o} \cong \bar{o}/\bar{p}$ the elements $\bar{\alpha}_0, \bar{\alpha}_1, \cdots, \bar{\alpha}_{n-1}$ are mapped upon elements of $K_p (= o/p)$. Since $K(\vartheta_1) = \Delta_p = K_p \cap \bar{\Delta}$, the elements $1, \theta, \cdots, \theta^{n-1}$ are linearly independent not only over $K(\vartheta_1)$, but also over K_p (Lemma 1). Consequently $\bar{\alpha}$ cannot belong to \bar{p} , unless all the elements $\bar{\alpha}_0, \bar{\alpha}_1, \cdots, \bar{\alpha}_{n-1}$ belong to \bar{p} . We have: $\bar{\alpha}_i \equiv \alpha_{i0} + \alpha_{i1}\omega + \cdots + \alpha_{i,m-1}\omega^{m-1}(o^*(\omega - \vartheta_1))$,

¹⁰ By $K_p \cap \bar{\Delta}$ is meant the intersection of $\bar{\Delta}$ (normal finite extension of K) with the relative algebraic closure K'_p of K in K_p , in the same sense as Δ was defined by the relation: $\Delta = K^* \cap K'$. See Section 1.

and if $\tilde{\alpha}_i \equiv 0(\tilde{p})$, then $\alpha_{i0} + \alpha_{i1}\omega + \dots + \alpha_{i,m-1}\omega^{m-1} \equiv 0(p)$, since $p = \tilde{p} \wedge o$. This shows that if $\tilde{\alpha}_i \equiv 0(\tilde{p})$, then $\tilde{\alpha}_i \equiv 0(\tilde{o}p, \omega - \vartheta_1)$, and consequently also $\tilde{\alpha} \equiv 0(\tilde{o}p, \omega - \vartheta_1)$, which proves the relation (6).

From (6) it follows already that the number of prime \tilde{o} -ideals which lie over p is finite, since it cannot be greater than $m (= [\Delta_p: K])$. Let these prime ideals be $\tilde{p}_1, \dots, \tilde{p}_h$, and let

$$(6') \quad \tilde{p}_i = (\tilde{o}p, \omega - \vartheta_i), \quad (i = 1, 2, \dots, h, h \leq m).$$

Since the \tilde{p}_i are also maximal, we have

$$(7) \quad [\tilde{p}_1, \dots, \tilde{p}_h] = \tilde{p}_1 \cdots \tilde{p}_h = (\tilde{o}p, \prod_{i=1}^h (\omega - \vartheta_i)).$$

Let $\xi = \prod_{i=h+1}^m (\omega - \vartheta_i)$ and let us consider the ideal $(\tilde{o}p, \xi)$. We assert that it is the unit ideal. Namely, in the contrary case let \tilde{p} be a prime ideal divisor of $(\tilde{o}p, \xi)$. Since $p \subset \tilde{p}$ and p is maximal, \tilde{p} must lie over p . Hence \tilde{p} must be one of the ideals $\tilde{p}_1, \dots, \tilde{p}_h$, say $\tilde{p} = \tilde{p}_1$. Now since $\xi \equiv 0(\tilde{p})$, one of the factors $\omega - \vartheta_i$, $i = h+1, \dots, m$, must belong to \tilde{p}_1 , say $\omega - \vartheta_{h+1} \equiv 0(\tilde{p}_1)$. Hence $\vartheta_1 - \vartheta_{h+1} \equiv 0(\tilde{p}_1)$, and this is impossible, since $\vartheta_1 \neq \vartheta_{h+1}$ and since $\vartheta_1 - \vartheta_{h+1}$ is an element of the subfield $\tilde{\Delta}$ of \tilde{o} .

It is therefore proved that $(\tilde{o}p, \xi) = \tilde{o}$. Consequently $\prod_{i=1}^h (\omega - \vartheta_i) \equiv 0(\tilde{o}p)$, since $\tilde{\xi} \cdot \prod_{i=1}^h (\omega - \vartheta_i) = f(\omega) \equiv 0(p)$. Comparing with (7) we find:

$$[\tilde{p}_1, \dots, \tilde{p}_h] = \tilde{o}p,$$

as was asserted.

5. It can be shown by examples that Theorem 2 is not generally true for non-maximal ideals.¹¹ For arbitrary prime o -ideals some weaker result

¹¹ Let K be the field of rational numbers, and let $\Sigma = K(\sqrt{2})(x, y)$, where x, y are independent variables. We put $K^* = K(\sqrt{2})$, $o = K[x, y, z]$, where $z = \sqrt{2} \cdot xy$. We have $\Sigma^* = K^*\Sigma = \Sigma$ and $o^* = K(\sqrt{2})[x, y]$. Let $p = o \cdot (y^2 - 2, z - 2x)$. Observing that every element of o can be put in the form $f(x, y) + z \cdot g(x, y)$, where $f(x, y), g(x, y) \in K[x, y]$, it is a straightforward matter to verify that p is prime. It is not maximal, since it is contained in the prime ideal $o(x, z, y^2 - 2)$. We have $o^*p = o^*(y^2 - 2, \sqrt{2} \cdot x(y - \sqrt{2})) = [p^*, p^*_o]$, where

$$p^* = o^*(y - \sqrt{2}), \quad p^*_o = o^*(x, y + \sqrt{2}).$$

The ideal p^* lies over p . In fact, any element $f(x, y) + zg(x, y)$, reduced modulo p , gives a residue of the form $A(x) + yB(x)$ (since $y^2 \equiv 2(p)$ and $z \equiv 2x(p)$). Here $A(x)$ and $B(x)$ are in $K[x]$. Should this residue belong to p^* , it is necessary that $A(x) + \sqrt{2} \cdot B(x)$ be identically zero. Hence $A(x) + yB(x)$ is also identically zero, and this shows that $p^* \wedge o = p$. However, the ideal p^*_o lies over the prime ideal $o \cdot (x, z, y^2 - 2)$ which is a proper divisor of p . It is remarkable that in this example o^*p possesses even an isolated component p^*_o different from p^* , since $p^* \not\equiv 0(p^*_o)$.

can be established by the usual artifice of quotient rings. Let \mathfrak{p} be an arbitrary prime \mathfrak{o} -ideal and let $\mathfrak{Z} = \mathfrak{o}_{\mathfrak{p}}$ be the quotient ring of \mathfrak{p} .¹² Let $\mathfrak{Z}^* = \mathfrak{K}^*\mathfrak{Z}$. It is well known that the prime ideals in \mathfrak{Z} correspond in one to one fashion to \mathfrak{p} and the prime \mathfrak{o} -ideals which are contained in \mathfrak{p} . If \mathfrak{p}_1 and \mathfrak{P}_1 are corresponding prime ideals in \mathfrak{o} and \mathfrak{Z} respectively, then $\mathfrak{P}_1 = \mathfrak{Z} \cdot \mathfrak{p}_1$, $\mathfrak{p}_1 = \mathfrak{P}_1 \cap \mathfrak{o}$. The prime ideal $\mathfrak{P} = \mathfrak{Z} \cdot \mathfrak{p}$ is maximal in \mathfrak{Z} . By Theorem 2 we have therefore:

$$(8) \quad \mathfrak{Z}^*\mathfrak{P} = [\mathfrak{P}_1^*, \mathfrak{P}_2^*, \dots],$$

where $\mathfrak{P}_1^*, \mathfrak{P}_2^*, \dots$ are the prime ideals in \mathfrak{Z}^* which lie over \mathfrak{P} .

Similarly, it can be shown in a simple manner that the prime ideals in \mathfrak{Z}^* correspond, one to one, to the prime ideals in \mathfrak{o}^* which lie over \mathfrak{p} or over prime multiples of \mathfrak{p} . The correspondence is again the one of contracted and extended ideals.¹³ Let $\mathfrak{P}_i^* \cap \mathfrak{o}^* = \mathfrak{p}_i^*$ and put

$$\begin{aligned} \mathfrak{o}^*\mathfrak{p} &= \mathfrak{m}^* \\ [\mathfrak{p}_1^*, \mathfrak{p}_2^*, \dots] &= \mathfrak{m}_1^*. \end{aligned}$$

We have evidently: $\mathfrak{Z}^*\mathfrak{m}^* = \mathfrak{Z}^*\mathfrak{m}_1^* = [\mathfrak{P}_1^*, \mathfrak{P}_2^*, \dots]$. Let us assume that Hilbert's basis theorem holds in \mathfrak{o}^* . From the relation $\mathfrak{Z}^*\mathfrak{m}^* = \mathfrak{Z}^*\mathfrak{m}_1^*$ follows that for every $\alpha^* \subset \mathfrak{m}_1^*$ there exists an element α in \mathfrak{o} but not in \mathfrak{p} , such that $\alpha^*\alpha \subset \mathfrak{m}^*$. By Hilbert's basis theorem there exists then an element β in \mathfrak{o} , not in \mathfrak{p} , such that $\beta\mathfrak{m}_1^* \equiv 0(\mathfrak{m}^*)$. This shows that $\mathfrak{p}_1^*, \mathfrak{p}_2^*, \dots$ are isolated components of \mathfrak{m}^* . Since $\mathfrak{o}^*\mathfrak{p} \cap \mathfrak{o} = \mathfrak{p}$, by Theorem 1, it follows that the decomposition of $\mathfrak{o}^*\mathfrak{p}$ into primary components is of the form

$$\mathfrak{o}^*\mathfrak{p} = [\mathfrak{p}_1^*, \mathfrak{p}_2^*, \dots; \mathfrak{q}'_1, \mathfrak{q}'_2, \dots],$$

where the prime ideals $\mathfrak{p}_1^*, \mathfrak{p}_2^*, \dots$ to which $\mathfrak{q}'_1, \mathfrak{q}'_2, \dots$ belong, lie over proper prime divisors of \mathfrak{p} .

6. The following theorem, which we shall have occasion to use in the sequel, gives a sufficient condition that $\mathfrak{o}^*\mathfrak{p}$ be prime, where \mathfrak{p} is now an arbitrary prime \mathfrak{o} -ideal, maximal or not.

¹² $\mathfrak{o}_{\mathfrak{p}}$ consists of all quotients α/β , $\alpha, \beta \subset \mathfrak{o}$, $\beta \not\equiv 0(\mathfrak{p})$.

¹³ The elements of \mathfrak{Z}^* are all of the form α^*/α , $\alpha^* \subset \mathfrak{o}^*$, $\alpha \subset \mathfrak{o}$, $\alpha \not\equiv 0(\mathfrak{p})$. Let \mathfrak{p}^* be a prime \mathfrak{o}^* -ideal which lies over a prime multiple of \mathfrak{p} , and let $\mathfrak{P}^* = \mathfrak{Z}^*\mathfrak{p}^*$. Let α^*/α , β^*/β be two elements in \mathfrak{Z}^* whose product is in \mathfrak{P}^* . Then $\alpha^*\beta^*/\alpha\beta = \gamma^*/\gamma$, where $\gamma^* \equiv 0(\mathfrak{p}^*)$, and therefore $\gamma\alpha^*\beta^* \equiv 0(\mathfrak{p}^*)$. Since $\mathfrak{p}^* \cap \mathfrak{o} \equiv 0(\mathfrak{p})$, it follows that $\gamma \not\equiv 0(\mathfrak{p}^*)$, and hence either α^* or β^* is in \mathfrak{p}^* , i.e. either α^*/α or β^*/β is in \mathfrak{P}^* . This shows that \mathfrak{P}^* is prime. Let $\alpha^* \subset \mathfrak{P}^* \cap \mathfrak{o}^*$, $\alpha^* = \beta^*/\beta$, $\beta^* \equiv 0(\mathfrak{p}^*)$. Then $\alpha^*\beta \equiv 0(\mathfrak{p}^*)$, and it follows by the same argument that α^* is in \mathfrak{p}^* . This shows that $\mathfrak{P}^* \cap \mathfrak{o}^* = \mathfrak{p}^*$.

If $\mathfrak{p}^* \cap \mathfrak{o} \not\equiv 0(\mathfrak{p})$, then let α be an element of \mathfrak{o} which is in \mathfrak{p}^* but not in \mathfrak{p} . Since α is a unit in \mathfrak{Z}^* , it follows that $\mathfrak{Z}^*\mathfrak{p}^* = \mathfrak{Z}^*$.

If \mathfrak{P}^* is an arbitrary prime ideal in \mathfrak{Z}^* , and if $\mathfrak{p}^* = \mathfrak{P}^* \cap \mathfrak{o}^*$, then any element α^*/α in \mathfrak{P}^* is such that α^* is in \mathfrak{p}^* , since α is a unit in \mathfrak{Z}^* . This shows that $\mathfrak{P}^* = \mathfrak{Z}^*\mathfrak{p}^*$.

THEOREM 3. *A sufficient condition that $\mathfrak{o}^* \mathfrak{p}$ be prime is that Δ' be the intersection of the fields $K_{\mathfrak{p}}$ and K^* .*

Proof. Let \mathfrak{p}^* be a prime \mathfrak{o}^* -ideal over \mathfrak{p} . Every element α^* of \mathfrak{o}^* can be written in the form: $\alpha^* = \alpha_0 + \alpha_1\theta + \cdots + \alpha_{g-1}\theta^{g-1}$, where $\alpha_i \in \mathfrak{o}$ and θ is an element of K^* , of degree g over Δ' . If $\alpha^* = 0(\mathfrak{p}^*)$, then passing to the residue field $K_{\mathfrak{p}^*}$, we find the relation: $\bar{\alpha}_0 + \bar{\alpha}_1\theta + \cdots + \bar{\alpha}_{g-1}\theta^{g-1} = 0$, where $\bar{\alpha}_i \in K_{\mathfrak{p}}$. Now if $K^* \cap K_{\mathfrak{p}} = \Delta'$, then, by Lemma 1, the elements $1, \theta, \cdots, \theta^{g-1}$ are linearly independent over $K_{\mathfrak{p}}$. Hence $\bar{\alpha}_i = 0$, i. e. $\alpha_i \equiv 0(\mathfrak{p})$. This shows that α^* is contained in $\mathfrak{o}^* \mathfrak{p}$, consequently $\mathfrak{o}^* \mathfrak{p} = \mathfrak{p}^*$, q. e. d.

Let K_0 be the largest subfield of $K_{\mathfrak{p}}$ which is algebraic over K , i. e. K_0 is the relative algebraic closure of K in $K_{\mathfrak{p}}$. Let K^* be the least normal extension of K which contains K_0 and let $\mathfrak{o}^* = K^*_{\mathfrak{o}}$. If \mathfrak{p}^* is any prime \mathfrak{o}^* -ideal over \mathfrak{p} , then, by a result proved in Section 2, we have: $K^*_{\mathfrak{p}^*} = K^* \cdot K_{\mathfrak{p}}$. In view of our choice of K^* , it follows that this field is algebraically closed in $K^*_{\mathfrak{p}^*}$. Hence if K^*_1 is any normal extension of the new ground field K^* , the condition of Theorem 3 is satisfied and \mathfrak{p}^* will remain prime when we pass from \mathfrak{o}^* to the ring $K^*_1 \mathfrak{o}^*$. This is true, in particular, if we pass from K^* to the algebraically closed field determined by K . In other words: *the extension $K \rightarrow K^*$ causes the maximal splitting of \mathfrak{p} into prime ideals.*

II. Algebraic varieties over arbitrary ground fields.

7. Let Σ be a field of algebraic functions of r independent variables, over an arbitrary ground field K of characteristic zero.¹⁴ We do not assume that K is algebraically closed in Σ . Let ξ_1, \cdots, ξ_n be a set of generators of Σ , i. e. $\Sigma = K(\xi_1, \cdots, \xi_n)$, and let $\mathfrak{o} = K[\xi_1, \cdots, \xi_n]$ be the ring consisting of those elements of Σ which can be expressed as polynomials in ξ_1, \cdots, ξ_n . With the elements ξ_i we associate an irreducible algebraic r -dimensional variety V_r whose general point has coördinates ξ_1, \cdots, ξ_n . A point P of V_r shall be associated with a prime zero-dimensional ideal \mathfrak{p}_0 in \mathfrak{o} . The geometric terms: "variety," "coördinates," "point," are so far purely formal and conventional expressions. To confer upon these terms a geometric reality it is necessary to imbed our V_r in an affine n -dimensional space S_n^{Λ} over some field Λ .¹⁵ The field Λ may be either K itself or an algebraic extension of K . Now the residue class field $K_{\mathfrak{p}_0}$ ($= \mathfrak{o}/\mathfrak{p}_0$) of the prime zero-dimensional ideal \mathfrak{p}_0 may very well be a proper extension of K (necessarily algebraic). Hence, in general, there

¹⁴ We assume, of course, that Σ is a finite extension of K (of degree of transcendency r).

¹⁵ By the symbol S_n^{Λ} we mean an affine n -space in which every point has coördinates a_1, \cdots, a_n , $a_i \in \Lambda$.

will not exist elements c_1, \dots, c_n in K such that $\xi_i \equiv c_i(p_0)$. In such a case our point P is not represented by a geometric point of S_n^K . On the other hand, if we take for Λ some normal extension of K , for instance the algebraically closed field determined by K , then the results of the preceding sections show that P may be represented in S_n^Λ by a set of points.¹⁶

More generally, we associate with a prime s -dimensional ideal p_s in \mathfrak{o} an irreducible algebraic s -dimensional subvariety V_s of V_r . The coordinates of the general point of this V_s are the elements ξ_1, \dots, ξ_n upon which the elements ξ_1, \dots, ξ_n are mapped in the homomorphism $\mathfrak{o} \cong \mathfrak{o}/p_s$. The residue class field K_{p_s} (i. e. the quotient field of \mathfrak{o}/p_s) is the field of rational functions on $V_s: K_{p_s} = K(\xi_1, \dots, \xi_n)$. Given two irreducible subvarieties V_s and V'_σ of V_r , defined by the prime ideals p_s and p'_σ respectively, we say that V_s belongs to V'_σ if $p'_\sigma \equiv 0(p_s)$.

8. Let V_s be an irreducible s -dimensional subvariety of V_r and let $p = p_s$ be the corresponding prime s -dimensional \mathfrak{o} -ideal. We consider the quotient ring $\mathfrak{S} = \mathfrak{o}_p$. The ideal $\mathfrak{S} \cdot p = \mathfrak{P}$ is prime and maximal in \mathfrak{S} and we have $\mathfrak{P} \wedge \mathfrak{o} = p$.¹⁷ The quotient ring $\mathfrak{S}_{\mathfrak{P}}$ is evidently \mathfrak{S} itself, and the residue class field of \mathfrak{P} ($= \mathfrak{S}/\mathfrak{P}$) coincides with K_p .

DEFINITION. V_s is said to be a simple subvariety of V_r if there exist $r-s$ elements $\eta_1, \dots, \eta_{r-s}$ in \mathfrak{S} such that:

$$(9) \quad \mathfrak{S} \cdot (\eta_1, \dots, \eta_{r-s}) = \mathfrak{P}.$$

Elements such as $\eta_1, \dots, \eta_{r-s}$ shall be referred to in the sequel as uniformizing parameters along V_s , or at V_s .

We shall see later that if V_s is simple, then the uniformizing parameters η_i can already be found in the ring \mathfrak{o} . Now if $\eta_1, \dots, \eta_{r-s}$ are in \mathfrak{o} , then (9) is equivalent to the condition that p itself occur among the maximal primary components of the ideal $\mathfrak{o} \cdot (\eta_1, \dots, \eta_{r-s})$,¹⁷ i. e.

$$(9') \quad \mathfrak{o} \cdot (\eta_1, \dots, \eta_{r-s}) = [p, \dots],$$

where the right-hand side is a decomposition of $\mathfrak{o} \cdot (\eta_1, \dots, \eta_{r-s})$ into maximal primary components.

¹⁶ This set is finite since the relative algebraic closure of K in K_{p_0} is a finite extension of K . See the footnote ¹⁴ and the considerations at the end of Section 6.

¹⁷ Concerning the relationship between the prime ideals in \mathfrak{o} and in \mathfrak{S} see Section 5. To that we add that, more generally, there is a (1,1) correspondence between the \mathfrak{S} -ideal \mathfrak{A} and those \mathfrak{o} -ideals α which have the property that each maximal primary component of α is a multiple of p . If \mathfrak{A} and α are corresponding ideals, then $\mathfrak{A} = \mathfrak{S} \cdot \alpha$, $\alpha = \mathfrak{A} \wedge \mathfrak{o}$. If α is an arbitrary ideal in \mathfrak{o} , then the ideal $\mathfrak{o} \wedge \mathfrak{S} \cdot \alpha$ differs from α only by primary components which are not multiples of p (such primary components are missing in the decomposition of $\mathfrak{o} \wedge \mathfrak{S} \cdot \alpha$).

Our purpose is to derive from the above definition a number of characteristic properties of simple subvarieties. The results will be on the main generalization of theorems proved by us elsewhere¹⁸ for simple points in the case of an algebraically closed ground field K . Practically all the rest of this paper deals with simple points. Once the results for simple points are established, the extension of these results to simple subvarieties of any dimension is rapidly achieved by the usual artifice of a transcendental extension of the ground field.

Dealing with a point P of V_r , given by a prime zero-dimensional \mathfrak{o} -ideal \mathfrak{p} , we shall proceed in the following manner. The residue class field $K_{\mathfrak{p}}$ is a finite algebraic extension of K . Let K^* be the least normal extension of K which contains $K_{\mathfrak{p}}$. We take K^* as new ground field and we pass to the field $\Sigma^* = K^*\Sigma$ and to the ring $\mathfrak{o}^* = K^*\mathfrak{o} = K^*[\xi_1, \dots, \xi_n]$. Regarded as elements of Σ^* the elements ξ_1, \dots, ξ_n are the coördinates of the general point of an irreducible variety V_r^* . Let $\mathfrak{p}_1^*, \dots, \mathfrak{p}_h^*$ be the prime \mathfrak{o}^* -ideals which lie over \mathfrak{p} and let P_1^*, \dots, P_h^* be the corresponding points of V_r^* . We may say that these points P_i^* correspond to the point P , and that P splits into the h points P_i^* of V_r^* . By Lemma 2 (section 2), the residue class field $K_{\mathfrak{p}_i^*}^*$ coincides with $K^*K_{\mathfrak{p}}$, and since $K_{\mathfrak{p}} \subseteq K^*$, it follows that $K_{\mathfrak{p}_i^*}^* = K^*$. Thus on the new variety V_r^* we now are dealing with points P_i^* ($i = 1, 2, \dots, h$) which have the property that for each of them the residue class field $K_{\mathfrak{p}_i^*}^*$ coincides with the ground field K^* . If we now pass from K^* to the algebraically closed field determined by K , then each prime ideal \mathfrak{p}_i^* remains prime (section 6), and it stands to reason that the results valid in the case of algebraically closed ground fields¹⁸ can therefore be carried over to the points P_i^* of V_r^* . For this reason we study first the special case in which $K_{\mathfrak{p}} = K$. When this special case has been settled, the only thing left to do in the general case will be to study the finite ground field extension $K \rightarrow K^*$, where K^* is the least normal extension of K which contains $K_{\mathfrak{p}}$.

III. Simple points. Case $K_{\mathfrak{p}} = K$.

9. Let \mathfrak{p} be a prime zero-dimensional ideal in \mathfrak{o} ($= K[\xi_1, \dots, \xi_n]$) and let the corresponding point P of V_r be a simple point, with η_1, \dots, η_r as uniformizing parameters, and such that the residue class field $K_{\mathfrak{p}}$ ($= \mathfrak{o}/\mathfrak{p}$, since \mathfrak{p} is divisorless) coincides with the ground field K . Let K^* be the algebraically closed field determined by K , and let $\Sigma^* = K^*\Sigma$, $\mathfrak{o}^* = K^*\mathfrak{o} = K^*[\xi_1, \dots, \xi_n]$, $\mathfrak{S}^* = K^*\mathfrak{S}$, where $\mathfrak{S} = \mathfrak{o}_{\mathfrak{p}}$. As was pointed out in the preceding section, the

¹⁸ "Some results in the arithmetic theory of algebraic varieties," *American Journal of Mathematics*, vol. 61 (April, 1939), no. 2, pp. 249-294.

ideal $\mathfrak{o}^* \mathfrak{p} = \mathfrak{p}^*$ is prime and lies over \mathfrak{p} . It is maximal in \mathfrak{o}^* , hence is zero-dimensional, and defines a point P^* of the variety V_r^* .

LEMMA 3. $\mathfrak{Z}^* = \mathfrak{o}^*_{\mathfrak{p}^*}$.

Proof. Since the relation $\mathfrak{Z}^* \subseteq \mathfrak{o}^*_{\mathfrak{p}^*}$ is trivial, in view of the relation $\mathfrak{p}^* \cap \mathfrak{o} = \mathfrak{p}$, we have only to show that $\mathfrak{o}^*_{\mathfrak{p}^*}$ is contained in \mathfrak{Z}^* . Let α^*/β^* be an element of $\mathfrak{o}^*_{\mathfrak{p}^*}$, $\alpha^*, \beta^* \in \mathfrak{o}^*$, $\beta^* \not\equiv 0(\mathfrak{p}^*)$. Since \mathfrak{p}^* is maximal, the ideal $(\mathfrak{o}^* \mathfrak{p}, \beta^*)$ is the unit ideal, i. e. we have a relation of the form

$$(10) \quad A^* \beta^* = 1 + \xi^*.$$

Here ξ^* is an element of $\mathfrak{o}^* \mathfrak{p}$ and hence can be put in the form:

$$\xi^* = \xi_0 + \xi_1 \theta_1 + \cdots + \xi_{g-1} \theta_1^{g-1},$$

where $\xi_i \equiv 0(\mathfrak{p})$ and θ_1 is an element of K^* satisfying an irreducible equation of degree g over K . If $\theta_2, \dots, \theta_g$ are the conjugates of θ_1 over K and if we multiply (10) by $\prod_{i=2}^g (1 + \xi^*_i)$, where $\xi^*_i = \xi_0 + \xi_1 \theta_i + \cdots + \xi_{g-1} \theta_i^{g-1}$, we get a relation of the form:

$$B^* \beta^* = 1 + \eta,$$

where $\eta \equiv 0(\mathfrak{p})$ and $B^* \in \mathfrak{o}^*$. Hence $\alpha^*/\beta^* = \alpha^* B^* / (1 + \eta)$, and since $1 + \eta$ is an element of \mathfrak{o} , not in \mathfrak{p} , it follows that α^*/β^* belongs to \mathfrak{Z}^* , q. e. d.

Let $\mathfrak{P}^* = \mathfrak{Z}^* \mathfrak{p} = \mathfrak{Z}^* \mathfrak{P} = \mathfrak{Z}^* \mathfrak{p}^*$, where $\mathfrak{P} = \mathfrak{p} \cdot \mathfrak{o}_{\mathfrak{p}}$. By (9), we have

$$(11) \quad \mathfrak{Z}^*(\eta_1, \dots, \eta_r) = \mathfrak{P}^*,$$

and since by the preceding Lemma the quotient ring $\mathfrak{o}^*_{\mathfrak{p}^*}$ coincides with \mathfrak{Z}^* , it follows that P^* is a simple point of V_r^* , and that η_1, \dots, η_r are uniformizing parameters at P^* . Since K^* is algebraically closed, we are in position to apply the results of our paper.¹⁸

Since $K_{\mathfrak{p}} = K$, every element ω of \mathfrak{Z} satisfies a congruence of the form: $\omega \equiv c(\mathfrak{P})$, $c \in K$. In particular, let $\xi_i \equiv c_i(\mathfrak{P})$, $c_1, \dots, c_n \in K$. The point P is therefore represented by an actual point (c_1, \dots, c_n) of the affine S_n^K . We shall assume from now on that P is the origin of coördinates in S_n^K , whence $\xi_i \equiv 0(\mathfrak{P})$, $i = 1, 2, \dots, n$.

By (9), every element ω of \mathfrak{P} can be put in the form: $\omega = A_1 \eta_1 + \cdots + A_r \eta_r$, $A_i \in \mathfrak{Z}$. Let $A_i \equiv c_i(\mathfrak{P})$. Then

$$(12) \quad \omega \equiv c_1 \eta_1 + \cdots + c_r \eta_r (\mathfrak{P}^2).$$

A congruence such as (12) holds true for any element ω in \mathfrak{P} . Since $\mathfrak{P}^* = \mathfrak{Z}^* \mathfrak{P}$, it follows from (12) that $\omega \equiv c_1 \eta_1 + \cdots + c_r \eta_r (\mathfrak{P}^{*2})$. We have proved in ¹⁸ that in this last congruence the coefficients c_1, \dots, c_r are uniquely determined. From this we conclude immediately with the following:

THEOREM 4. *The coefficients c_1, \dots, c_r in (12) are uniquely determined and belong to K . The elements η_1, \dots, η_r are linearly independent mod \mathfrak{P}^{*2} over K^* . Moreover, we have the following relation:*

$$(13) \quad \mathfrak{P}^{*2} \wedge \mathfrak{S} = \mathfrak{P}^2.$$

By a similar argument and observing that $\mathfrak{P}^2 = \mathfrak{S}(\eta_1^2, \eta_1\eta_2, \dots, \eta_r^2)$, we find that any element ω in \mathfrak{P} satisfies a congruence of the form:

$$\omega = c_1\eta_1 + \dots + c_r\eta_r + c_{11}\eta_1^2 + c_{12}\eta_1\eta_2 + \dots + c_{rr}\eta_r^2 (\mathfrak{P}^3),$$

where the coefficients c_1, \dots, c_r are the same as in (12). Proceeding in the same fashion, we find, more generally, that for any element ω in \mathfrak{S} there exists a formal integral power series:

$$\psi_0 + \psi_1 + \dots + \psi_m + \dots,$$

where ψ_i is a form of degree i in η_1, \dots, η_r , with coefficients in K , such that

$$(14) \quad \omega \equiv \psi_0 + \psi_1 + \dots + \psi_m (\mathfrak{P}^{m+1}), \quad m\text{-arbitrary}.$$

Here $\psi_0 = 0$, if and only if $\omega \equiv 0 (\mathfrak{P})$. From (14) it follows that $\omega \equiv \psi_0 + \psi_1 + \dots + \psi_m (\mathfrak{P}^{*m+1})$, and we know that in this congruence the polynomial $\psi_0 + \psi_1 + \dots + \psi_m$ is uniquely determined by m .¹⁸ Hence, if $\omega \equiv 0 (\mathfrak{P}^{*m+1})$, then this polynomial must be identically zero, and consequently

$$(15) \quad \mathfrak{P}^{*m} \wedge \mathfrak{S} = \mathfrak{P}^m, \quad m\text{-arbitrary},$$

a generalization of (13).

The result to the effect that the uniformizing parameters at P are also uniformizing parameters at P^* , can be inverted. We show namely that if P is a simple point and if r elements $\omega_1, \dots, \omega_r$ in \mathfrak{o} are uniformizing parameters at P^* , then they are also uniformizing parameters at P , i. e. $\mathfrak{S}^*(\omega_1, \dots, \omega_r) = \mathfrak{P}^*$ implies $\mathfrak{S}(\omega_1, \dots, \omega_r) = \mathfrak{P}$. Let

$$(16) \quad \omega_i \equiv c_{i1}\eta_1 + \dots + c_{ir}\eta_r (\mathfrak{P}^{*2}),$$

$c_{ij} \in K$. It has been proved (¹⁸, Theorem 1) that the non-vanishing of the determinant $|c_{ij}|$ is a necessary and sufficient condition in order that $\omega_1, \dots, \omega_r$ be uniformizing parameters at P^* . Hence $|c_{ij}| \neq 0$, and since the c_{ij} are in K we conclude from (16) and (13) that η_1, \dots, η_r satisfy congruences of the form:

$$\eta_i \equiv d_{i1}\omega_1 + \dots + d_{ir}\omega_r (\mathfrak{P}^2), \quad (i = 1, 2, \dots, r)$$

$$d_{ij} \in K.$$

Hence, by (12), every element ω in \mathfrak{P} satisfies a congruence of the form: $\omega \equiv e_1\omega_1 + \dots + e_r\omega_r (\mathfrak{P}^2)$, $e_i \in K$. Denote the ideal $\mathfrak{S}(\omega_1, \dots, \omega_r)$ by \mathfrak{A} . The above relation implies the following relation:

$$(17) \quad (\mathfrak{A}, \mathfrak{P}^2) = \mathfrak{P}.$$

Since $\mathfrak{Z}^*\mathfrak{A} = \mathfrak{P}^*$, it follows that \mathfrak{A} has no prime ideal divisors in \mathfrak{Z} other than \mathfrak{P} .¹⁹ Consequently \mathfrak{A} is a primary ideal, with \mathfrak{P} as associated prime ideal, and this, in view of (17), implies that \mathfrak{A} coincides with \mathfrak{P} ,²⁰ as was asserted.

We now are in position to prove the following

THEOREM 5. *There exist uniformizing parameters $\omega_1, \dots, \omega_r$ at P which are elements of \mathfrak{o} and are such that \mathfrak{o} is integrally dependent on the ring $\mathbf{K}[\omega_1, \dots, \omega_r]$. Such parameters are furnished, for instance, by linear forms in ξ_1, \dots, ξ_n with non special coefficients in \mathbf{K} .*

Proof. Since the original uniformizing parameters η_1, \dots, η_r are polynomials in ξ_1, \dots, ξ_n , and since $\mathfrak{P} = \mathfrak{Z} \cdot (\xi_1, \dots, \xi_n)$, we have relations of the form: $\eta_i \equiv \sum_{j=1}^n c_{ij} \xi_j (\mathfrak{P}^2)$, $i = 1, 2, \dots, r$. The r forms $\sum_{j=1}^n c_{ij} \xi_j$ are linearly independent mod \mathfrak{P}^2 (Theorem 4). Hence if $\xi_i \equiv \sum_{j=1}^r e_{ij} \eta_j (\mathfrak{P}^2)$, then the n by r matrix (e_{ij}) must be of rank r . If then we put $\omega_i = \sum_{j=1}^n u_{ij} \xi_j$, $i = 1, 2, \dots, r$, then for non special constants u_{ij} in \mathbf{K} the r -row square matrix $(u_{ij}) (e_{jv})$ will be non singular. The elements $\omega_1, \dots, \omega_r$ will then be uniformizing parameters at P^* , hence also at P .

In addition, by a well known normalization theorem of E. Noether, for non special u_{ij} the ring \mathfrak{o} will be integrally dependent on $\mathbf{K}[\omega_1, \dots, \omega_r]$. This completes the proof of the theorem.

10. We have seen in the preceding section that if the point P is simple for V_r , then P^* is simple for V_r^* . It can be shown by examples that the converse is not generally true.²¹ We prove, however, the following

THEOREM 6. *Under the hypothesis $\mathbf{K}_{\mathfrak{p}} = \mathbf{K}$, a necessary and sufficient condition that P be a simple point of V_r is that P^* be a simple point of V_r^* and that \mathbf{K} be maximally algebraic in Σ (\mathbf{K} algebraically closed in Σ).*

Proof. The condition is sufficient. For if \mathbf{K} is maximally algebraic in Σ ,

¹⁹ Let \mathfrak{P}_1 be a prime ideal divisor of \mathfrak{A} . There exists in \mathfrak{Z}^* at least one prime ideal, say \mathfrak{P}_1^* , which lies over \mathfrak{P}_1 (Krull). Since \mathfrak{P}_1^* is a divisor of $\mathfrak{Z}^*\mathfrak{A}$, and since $\mathfrak{Z}^*\mathfrak{A} (= \mathfrak{P}^*)$ is maximal, necessarily $\mathfrak{P}_1^* = \mathfrak{P}^*$, whence $\mathfrak{P}_1 = \mathfrak{P}$.

²⁰ Let ρ be the exponent of \mathfrak{A} , i.e. let $\mathfrak{P}^\rho \equiv 0(\mathfrak{A})$, $\mathfrak{P}^{\rho-1} \not\equiv 0(\mathfrak{A})$. Assuming $\rho > 1$, we multiply (17) by $\mathfrak{P}^{\rho-2}$, getting $\mathfrak{P}^{\rho-1} = (\mathfrak{A} \cdot \mathfrak{P}^{\rho-2}, \mathfrak{P}^\rho) \equiv 0(\mathfrak{A})$, a contradiction.

²¹ We refer to the example given in the footnote³. Let $\mathfrak{p} = \mathfrak{o} \cdot (x, x\sqrt{2})$, $\mathfrak{p}^* = \mathfrak{o}^*(x)$. The point P^* is simple, and x is a uniformizing parameter at P^* . The quotient field $\mathbf{K}_{\mathfrak{p}}$ is obviously the field \mathbf{K} . However, P is not a simple point, since the ring $\mathfrak{p}/\mathfrak{p}^2$ is a \mathbf{K} -module of rank 2, while, according to Theorem 4, the ring $\mathfrak{p}/\mathfrak{p}^2$ for a simple point must be of rank r . In the present case we have $r = 1$.

then the relation $\mathfrak{Z}^*\mathfrak{M} \circ \mathfrak{Z} = \mathfrak{M}$ holds true for any \mathfrak{Z} -ideal \mathfrak{M} (Theorem 1). If, in addition, P^* is simple for V_r^* , then from the proof of Theorem 5 it follows that we can find uniformizing parameters $\omega_1, \dots, \omega_r$ at P^* which are elements of \mathfrak{Z} . We will have then the relation: $\mathfrak{Z}^*(\omega_1, \dots, \omega_r) = \mathfrak{P}^*$. Since $\mathfrak{P}^* \circ \mathfrak{Z} = \mathfrak{P}$ and since, by Theorem 1, $\mathfrak{Z}^*(\omega_1, \dots, \omega_r) \circ \mathfrak{Z} = \mathfrak{Z}^*(\omega_1, \dots, \omega_r)$, it follows that $\mathfrak{Z}^*(\omega_1, \dots, \omega_r) = \mathfrak{P}$, whence P is a simple point of V_r .

The condition is necessary. Let η_1, \dots, η_r be uniformizing parameters at P and let θ be an element of Σ which is algebraically dependent on K . Let $\theta = \alpha/\beta$, $\alpha, \beta \in \mathfrak{Z}$, and let

$$\beta = \psi_\rho(\eta_1, \dots, \eta_r) + \psi_{\rho+1}(\eta_1, \dots, \eta_r) + \dots, \psi_\rho \neq 0,$$

be the power series expansion for β (see preceding section, especially congruence (14)). The coefficients of these power series are in K . Since $\alpha = \theta\beta$ and $\theta \in K^*$, the element α will have the following power series expansion: $\alpha = \theta\psi_\rho + \theta\psi_{\rho+1} + \dots$. Since the coefficient of the form $\theta\psi_\rho$ must also be elements of K , it follows that θ is an element of K . Hence K is algebraically closed in Σ , q. e. d.²²

Let η_1, \dots, η_r be r algebraically independent elements in \mathfrak{o} such that \mathfrak{o} is integrally dependent on the ring $K[\eta_1, \dots, \eta_r]$. Let ω be an element of \mathfrak{o} and let $G(\eta_1, \dots, \eta_r; z)$ be the norm of $z - \omega$ with respect to the field $K(\eta_1, \dots, \eta_r)$. Let moreover $\eta_i \equiv c_i(p)$, $c_i \in K$. In the case of an algebraically closed field K we have proved the following (¹⁸, Theorem 4): *a necessary and sufficient condition that P be a simple point and that $\eta_1 - c_1, \dots, \eta_r - c_r$ be uniformizing parameters at P , is that there should exist an element ω in \mathfrak{o} such that $G'_\omega(\eta_1, \dots, \eta_r; \omega) \not\equiv 0(p)$.* Using Theorem 6 we are now in position to extend this result to the case under consideration ($K_p = K$).

Assume that P is a simple point and that the elements $\eta_1 - c_1, \dots, \eta_r - c_r$ are uniformizing parameters at P . The elements $\eta_1 - c_1, \dots, \eta_r - c_r$ are then also uniformizing parameters at P^* , and \mathfrak{o}^* is integrally dependent on the ring $K^*[\eta_1, \dots, \eta_r]$. By the quoted theorem, proved for the algebraically closed field K^* , there exists in \mathfrak{o}^* an element ω such that $F'_\omega(\eta_1, \dots, \eta_r; \omega) \not\equiv 0(p^*)$, where $F(\eta_1, \dots, \eta_r; z)$ is the norm of $z - \omega$ with respect to the field $K^*(\eta_1, \dots, \eta_r)$. More specifically we have shown (¹⁸, p. 269) that we may put $\omega = v_1\xi_1 + \dots + v_n\xi_n$, $v_i \in K^*$, provided the coefficients v_i do not satisfy certain linear relations with coefficients in K^* . Hence, we may choose the v_i in K , and we may therefore assume that ω is an element of \mathfrak{o} . The relation $F'_\omega \not\equiv 0(p^*)$ implies at any rate that $F(\eta_1, \dots, \eta_r; z)$ is irreducible (over

²² An immediate corollary of Theorem 6 is the following: if K is not maximally algebraic in Σ then the residue class field K_p for any simple point P of V_r is necessarily a proper algebraic extension of K . This is a special case of a more general theorem (Theorem 9) proved in Section 14.

K^*) and that ω is a primitive element of $\Sigma^*/K^*(\eta_1, \dots, \eta_r)$. By Theorem 6, K is algebraically closed in Σ . Hence, by Lemma 1, the relative degrees $[\Sigma^*: K^*(\eta_1, \dots, \eta_r)]$, $[\Sigma: K(\eta_1, \dots, \eta_r)]$ are the same. Consequently $F(\eta_1, \dots, \eta_r; z)$ is also the norm of $z - \omega$ with respect to the field $K(\eta_1, \dots, \eta_r)$, and since $F'_\omega \not\equiv 0(p^*)$ implies $F'_\omega \not\equiv 0(p)$, it is thus proved that our condition is necessary.

Conversely, assume $G'_\omega(\eta_1, \dots, \eta_r; \omega) \not\equiv 0(p)$, and let $F(\eta_1, \dots, \eta_r; z)$ be the norm of $z - \omega$ with respect to the field $K^*(\eta_1, \dots, \eta_r)$. Clearly F either coincides with G or is a proper factor of G , according as the relative degree $[\Sigma^*: K^*(\eta_1, \dots, \eta_r)]$ coincides with or is less than the relative degree $[\Sigma: K(\eta_1, \dots, \eta_r)]$. In either case the relation $G'_\omega \not\equiv 0(p)$ implies the relation $F'_\omega \not\equiv 0(p^*)$, and hence P^* is a simple point, with $\eta_1 - c_1, \dots, \eta_r - c_r$ as uniformizing parameters. To prove that P is a simple point of V_r , we have only to show, according to Theorem 6, that K is maximally algebraic in Σ . Let K' be the relative algebraic closure of K in Σ and let $[K': K] = g$. Let θ_1 be a primitive element of K' over K , so that $K' = K(\theta_1)$. Since the relative degrees $[\Sigma: K'(\eta_1, \dots, \eta_r)]$ and $[\Sigma^*: K^*(\eta_1, \dots, \eta_r)]$ are the same (Lemma 1), $F(\eta_1, \dots, \eta_r; z)$ is the also norm of $z - \omega$ with respect to the field $K'(\eta_1, \dots, \eta_r)$. Hence F is a polynomial in η_1, \dots, η_r, z and θ_1 , with coefficients in K : $F = F(\eta_1, \dots, \eta_r; z; \theta_1)$. If $\theta_2, \dots, \theta_g$ are the conjugates of θ_1 over K , then

$$(18) \quad G(\eta_1, \dots, \eta_r; z) = \prod_{i=1}^g F(\eta_1, \dots, \eta_r; z; \theta_i).$$

Let $\omega \equiv c(p)$, $c \subset K$ (since $K_p = K$). If we reduce the equation $F(\eta_1, \dots, \eta_r; \omega; \theta_1) = 0$ modulo p^* , we get $F(c_1, \dots, c_r; c; \theta_1) = 0$. Hence also $F(c_1, \dots, c_r; c; \theta_i) = 0$, $i = 1, 2, \dots, g$, i. e.

$$(18') \quad F(\eta_1, \dots, \eta_r; \omega; \theta_i) \equiv 0(p^*), \quad (i = 1, 2, \dots, g).$$

Now if g were greater than 1, then it would follow from (18) and (18') that $G'_\omega \equiv 0(p^*)$, i. e. $G'_\omega \equiv 0(p)$, since $G'_\omega \subset \mathfrak{o}$, a contradiction. Hence $g = 1$, i. e. $K' = K$, q. e. d.

Remark. Let $\mathfrak{o} \cdot (\eta_1, \dots, \eta_r) = [p, q_1, q_2, \dots]$ be the decomposition of the ideal $\mathfrak{o} \cdot (\eta_1, \dots, \eta_r)$ into maximal primary components (see (9')), where we assume that η_1, \dots, η_r are uniformizing parameters at the simple point P and that \mathfrak{o} is integrally dependent on $K[\eta_1, \dots, \eta_r]$. This last condition implies that the above primary components are all zero-dimensional. Let $\omega \equiv c(p)$. For algebraically closed ground fields we have proved (¹⁸, Theorem 4), that the elements ω such that $G'_\omega \not\equiv 0(p)$ are characterized by the condition: $\omega \not\equiv c(p_i)$, $i = 1, 2, \dots$, where p_1, p_2, \dots are the prime ideals to which q_1, q_2, \dots belong respectively. It is clear that this result holds true

also in the present case where $K_p = K$. It is sufficient to take into account the relations: $o^*(\eta_1, \dots, \eta_r) = o^*p \cdot o^*q_1 \cdot o^*q_2 \cdot \dots = [o^*p, o^*q_1, o^*q_2, \dots]$.

IV. Simple points. General case.

11. Let P be a point of V_r , p the corresponding prime zero-dimensional ideal in o . Following the plan outlined in Section 8, we extend our ground field K as follows: we pass from K to the least normal extension K^* which contains the residue class field K_p . We put: $\Sigma^* = K^*\Sigma$, $o^* = K^*o$, $\mathfrak{Z}^* = K^*\mathfrak{Z}$, where $\mathfrak{Z} = o_p$. Since K^* is a finite extension of K , there is only a finite number of prime o^* -ideals which lie over p , say p^*_1, \dots, p^*_h , and we have, by Theorem 2:

$$(19) \quad o^*p = [p^*_1, \dots, p^*_h] = p^*_1 \cdot \dots \cdot p^*_h.$$

We denote by P^*_i the point of V^*_r defined by the prime ideal p^*_i . For the residue class field $K^*_{p^*_i}$ we have the relation (Lemma 2, Section 2): $K^*_{p^*_i} = K^* \cdot K_p = K^*$, since $K_p \subseteq K^*$. Hence the residue class field at each point P^*_i coincides with the new ground field K^* . If then P^*_i is a simple point, we have, as far as V^*_r is concerned, the situation studied in the preceding Part III.

We prove the relation $\mathfrak{Z}^* = o^*_{p^*_1} \wedge o^*_{p^*_2} \wedge \dots \wedge o^*_{p^*_h}$.

Proof. It is clear that \mathfrak{Z}^* is contained in each of the quotient rings $o^*_{p^*_i}$. Hence to prove the above relation we have only to show that if η^* is an element which belongs to each quotient ring $o^*_{p^*_i}$, then $\eta^* \in \mathfrak{Z}^*$. We can write $\eta^* = \alpha^*_i / \beta^*_i$, $i = 1, 2, \dots, h$, where $\alpha^*_i, \beta^*_i \in o^*$, $\beta^*_i \not\equiv 0 (p^*_i)$. Since the p^*_i are maximal ideals, we can find, for each $i = 1, 2, \dots, h$, an element ω^*_i in o^* satisfying the congruences: $\omega^*_i \equiv 1 (p^*_i)$, $\omega^*_i \equiv 0 (p^*_j)$, $j \neq i$. If we put $\gamma^* = \alpha^*_1 \omega^*_1 + \dots + \alpha^*_h \omega^*_h$, $\delta^* = \beta^*_1 \omega^*_1 + \dots + \beta^*_h \omega^*_h$, then $\eta^* = \gamma^* / \delta^*$ and $\delta^* \equiv \beta^*_i \not\equiv 0 (p^*_i)$. We have thus found for η^* a quotient representation γ^* / δ^* in which the denominator δ^* is not in any of the ideals p^*_i , $i = 1, 2, \dots, h$. But then the ideal $(o^*p, o^*\delta^*)$ is the unit ideal, and the rest of the proof is the same as that of Lemma 3 (Section 9).

It now follows immediately that \mathfrak{Z}^* has h and only h distinct prime zero-dimensional ideals $\mathfrak{P}_1, \mathfrak{P}_2, \dots, \mathfrak{P}_h$ which correspond to the ideals p^*_1, \dots, p^*_h respectively. Namely, let $\mathfrak{Z}^*_{p^*_i} = o^*_{p^*_i}$ and let \mathfrak{P}^*_i be the prime zero-dimensional ideal of $\mathfrak{Z}^*_{p^*_i}$:

$$(20) \quad \mathfrak{P}^*_i = \mathfrak{Z}^*_{p^*_i} \cdot p^*_i, \quad p^*_i = o^* \wedge \mathfrak{P}^*_i.$$

Then

$$(21) \quad \bar{\mathfrak{P}}_i = \mathfrak{P}^*_i \wedge \mathfrak{Z}^* = \mathfrak{Z}^* \cdot p^*_i,^{23} \quad (i = 1, 2, \dots, h).$$

²³ Let $\bar{\mathfrak{P}}$ be a prime zero-dimensional ideal in \mathfrak{Z}^* , and let $\bar{\mathfrak{P}} \wedge o^* = p^*$. The prime ideal p^* is zero-dimensional and must coincide with one of the ideals p^*_1, \dots, p^*_h .

The quotient ring $\mathfrak{S}^*/\mathfrak{P}_i$ contains the quotient ring $\mathfrak{S}^*_{\mathfrak{P}_i}$, since $\mathfrak{o}^* \wedge \mathfrak{P}_i = \mathfrak{p}^*_{\mathfrak{P}_i}$. On the other hand we have $\mathfrak{S}^*/\mathfrak{P}_i \subseteq \mathfrak{S}^*_{\mathfrak{P}_i}$, since $\mathfrak{P}^*_{\mathfrak{P}_i} \wedge \mathfrak{S}^* = \mathfrak{P}^*_{\mathfrak{P}_i}$. Hence

$$(22) \quad \mathfrak{S}^*/\mathfrak{P}_i = \mathfrak{S}^*_{\mathfrak{P}_i} = \mathfrak{o}^*_{\mathfrak{P}^*_{\mathfrak{P}_i}}.$$

12. The relations (20-22) are true for any point P , simple or not. Now we assume that P is a simple point and that η_1, \dots, η_r are uniformizing parameters at P . We have then $\mathfrak{S} \cdot (\eta_1, \dots, \eta_r) = \mathfrak{P} = \mathfrak{S} \cdot \mathfrak{p}$. Hence, by (19) and (21), $\mathfrak{S}^* \cdot (\eta_1, \dots, \eta_r) = \mathfrak{P}^*_1 \cdot \dots \cdot \mathfrak{P}^*_h$, and consequently, in view of (22), $\mathfrak{S}^*_{\mathfrak{P}_i} \cdot (\eta_1, \dots, \eta_r) = \mathfrak{P}^*_{\mathfrak{P}_i}$. This shows that each of the points $P^*_{\mathfrak{P}_i}$ is a simple point of $V^*_{\mathfrak{P}_i}$ and that η_1, \dots, η_r are uniformizing parameters at $P^*_{\mathfrak{P}_i}$. The following theorem is in a sense the converse:

THEOREM 7. *If P is a simple point of V_r and if the elements $\omega_1, \dots, \omega_r$ of \mathfrak{S} are uniformizing parameters at one of the points $P^*_{\mathfrak{P}_i}$, then they are also uniformizing parameters at P .*

Proof. For the proof, we first establish the following relation:

$$(23) \quad \mathfrak{P}^*_{\mathfrak{P}_i} \wedge \mathfrak{S}^m = \mathfrak{P}^m, \quad m - \text{an arbitrary integer} \geq 0.$$

Since $\mathfrak{P}^m \equiv 0(\mathfrak{P}^*_{\mathfrak{P}_i} \wedge \mathfrak{S}^m)$, we have only to show that any element α of $\mathfrak{P}^*_{\mathfrak{P}_i} \wedge \mathfrak{S}$ is contained in \mathfrak{P}^m . Let η_1, \dots, η_r be uniformizing parameters at P . The element α certainly belongs to \mathfrak{P} , and since $\mathfrak{P} = \mathfrak{S} \cdot (\eta_1, \dots, \eta_r)$, we have: $\alpha = A_1\eta_1 + \dots + A_r\eta_r$, $A_i \in \mathfrak{S}$. If A_1, \dots, A_r also belong to \mathfrak{P} , then we can put α in the form: $\alpha = \sum_{i,j} A_{ij}\eta_i\eta_j$. Continuing in this manner we will

ultimately get for α an expression of the form: $\alpha = \phi_s(\eta_1, \dots, \eta_r)$, where ϕ_s is a form of degree s in η_1, \dots, η_r whose coefficients $A_{(j)} (= A_{j_1 \dots j_r})$ are elements of \mathfrak{S} , and the following are the only two possibilities: (a) either $s \geq m$, or (b) $s < m$ and not all the elements $A_{(j)}$ are in \mathfrak{P} . In the case (a) we have $\alpha \equiv 0(\mathfrak{P}^m)$, as was asserted. We show that the case (b) leads to a contradiction. Let us denote by $\phi_s^{(0)} (= \phi_s^{(0)}(\eta_1, \dots, \eta_r))$ the reduced form obtained from $\phi_s(\eta_1, \dots, \eta_r)$ by reducing the elements $A_{(j)}$ modulo $\mathfrak{P}^*_{\mathfrak{P}_i}$ to elements of K^* .²⁴ By hypothesis the form $\phi_s^{(0)}$ is not identically zero in η_1, \dots, η_r . Since $\alpha \equiv 0(\mathfrak{P}^*_{\mathfrak{P}_i} \wedge \mathfrak{S}^m)$ and since $s < m$, we have obviously the congruence:

$$\phi_s^{(0)}(\eta_1, \dots, \eta_r) \equiv 0(\mathfrak{P}^*_{\mathfrak{P}_i} \wedge \mathfrak{S}^{s+1}).$$

because any element which is not in any one of the ideals $\mathfrak{p}^*_{\mathfrak{P}_1}, \dots, \mathfrak{p}^*_{\mathfrak{P}_h}$ is a unit in \mathfrak{S}^* . Let, say, $\mathfrak{P} \wedge \mathfrak{o}^* = \mathfrak{p}^*_{\mathfrak{P}_1}$. If α^*/β is an element of \mathfrak{P} , where $\alpha^* \in \mathfrak{o}^*, \beta \in \mathfrak{o}, \beta \not\equiv 0(\mathfrak{p})$, then $\alpha^* \equiv 0(\mathfrak{p}^*_{\mathfrak{P}_1})$, since β is a unit in \mathfrak{S}^* . Hence $\mathfrak{P} = \mathfrak{S}^* \cdot \mathfrak{p}^*_{\mathfrak{P}_1}$. \mathfrak{S}^* contains at least h distinct prime zero-dimensional ideals, namely the ideals $\mathfrak{P}^*_i = \mathfrak{P}^*_{\mathfrak{P}_i} \wedge \mathfrak{S}^*, i = 1, 2, \dots, h$. They are distinct, because $\mathfrak{P}^*_i \wedge \mathfrak{o}^* = \mathfrak{P}^*_{\mathfrak{P}_i} \wedge \mathfrak{o}^* = \mathfrak{p}^*_{\mathfrak{P}_i}$. Hence the h ideals $\mathfrak{S}^* \mathfrak{p}^*_{\mathfrak{P}_i}, i = 1, 2, \dots, h$, are the only prime zero-dimensional ideals in \mathfrak{S}^* .

²⁴We recall that K^* coincides with the residue class field of $\mathfrak{P}^*_{\mathfrak{P}_i}$.

This congruence is in contradiction with the uniqueness of the polynomial $\psi_0 + \psi_1 + \dots + \psi_m$ satisfying the congruence (14) (Section 9) (with \mathfrak{P} replaced by \mathfrak{P}^*_i). The uniqueness of this polynomial, for a given element ω , shows namely that no form of degree s in η_1, \dots, η_r , with coefficients in K^* , can belong to $\mathfrak{P}^{*i, m+1}$. The relation (23) is thus proved.

Now let $\omega_1, \dots, \omega_r$ be elements of \mathfrak{o} which are uniformizing parameters at one of the points P^*_i , say at P^*_1 . We have then the relation: $\mathfrak{S}^*_{1 \cdot}(\omega_1, \dots, \omega_r) = \mathfrak{P}^*_1$. Let α be any element of \mathfrak{P} and let, by (12) (Section 9):

$$(24) \quad \alpha \equiv c^*_{1\omega_1} + \dots + c^*_{r\omega_r}(\mathfrak{P}^{*1^2}), \quad c^*_{ij} \in K^*.$$

Let, in particular,

$$(25) \quad \eta_i \equiv c^*_{i1\omega_1} + \dots + c^*_{ir\omega_r}(\mathfrak{P}^{*1^2}), \quad c^*_{ij} \in K^*, \\ (i = 1, 2, \dots, r).$$

Since the η^*_i are uniformizing parameters at P , we have:

$$\omega_i = A_{i1}\eta_1 + \dots + A_{ir}\eta_r, \quad A_{ij} \in \mathfrak{S}.$$

Let $A_{ij} \equiv d^*_{ij}(\mathfrak{P}^*_1)$. Since the A_{ij} are in \mathfrak{S} , the elements d^*_{ij} are not only in K^* but also in K_p . We have:

$$(26) \quad \omega_i \equiv d^*_{i1}\eta_1 + \dots + d^*_{ir}\eta_r(\mathfrak{P}^{*1^2}), \quad (i = 1, 2, \dots, r).$$

The matrix (c^*_{ij}) in (25) is non-singular, since η_1, \dots, η_r are also uniformizing parameters at P^*_1 . Comparing with (26) we see that the matrix (c^*_{ij}) is the inverse of the matrix (d^*_{ij}) , and consequently, since $d^*_{ij} \in K_p$, we conclude that the c^*_{ij} also belong to K_p . Now, let $\alpha \equiv d^*_{1\eta_1} + \dots + d^*_{r\eta_r}(\mathfrak{P}^{*1^2})$. The same argument by which the d^*_{ij} have been proved to belong to K_p shows that $d^*_{1\eta_1}, \dots, d^*_{r\eta_r}$ belong to K_p . Since $c^*_{ij} = \sum_{j=1}^r d^*_{ij}c^*_{ji}$, it follows that also the coefficients c^*_{ij} in (24) belong to K_p . Consequently there exist in \mathfrak{S} elements A_1, \dots, A_r such that $A_i \equiv c^*_{i1}(\mathfrak{P}^*_1)$, and for such elements the relation (24) implies the following:

$$\alpha \equiv A_1\omega_1 + \dots + A_r\omega_r(\mathfrak{P}^{*1^2}).$$

Since $\alpha - A_1\omega_1 - \dots - A_r\omega_r \in \mathfrak{S}$, this last congruence, in view of (23), implies that:

$$\alpha \equiv A_1\omega_1 + \dots + A_r\omega_r(\mathfrak{P}^2), \quad A_i \in \mathfrak{S}.$$

Such a congruence holds for any element α in \mathfrak{P} . Consequently

$$(\mathfrak{S} \cdot (\omega_1, \dots, \omega_r), \mathfrak{P}^2) = \mathfrak{P},$$

and therefore, by an argument used before (footnotes ^{19, 20}),

$$\mathfrak{S} \cdot (\omega_1, \dots, \omega_r) = \mathfrak{P},$$

i. e. $\omega_1, \dots, \omega_r$ are uniformizing parameters at P , q. e. d.

13. Let $\xi_i^* = \sum_{j=1}^n u_{ij} \xi_j$, $i = 1, 2, \dots, r$, $u_{ij} \in K^*$, and let $\xi_i^* \equiv v_i^*(\mathfrak{P}^*_1)$, $v_i^* \in K^*$. By Theorem 5, the elements $\xi_i^* - v_i^*$ are uniformizing parameters at P^*_1 , provided the u_{ij} are "non special," and moreover, \mathfrak{o}^* is integrally dependent on $K^*[\xi_1^*, \dots, \xi_r^*]$. Since the values of the u_{ij} to be avoided are those which satisfy certain algebraic relations, we may choose the u_{ij} in K , and therefore we may assume that ξ_1^*, \dots, ξ_r^* are elements of \mathfrak{o} . The constants v_1^*, \dots, v_r^* are algebraic over K . Let $f_i(v_i^*) = 0$ be the irreducible equation over K which is satisfied by v_i^* , and let us consider the elements $\omega_i = f_i(\xi_i^*)$, $i = 1, 2, \dots, r$. These are elements in \mathfrak{o} and \mathfrak{o} is integrally dependent on the ring $K[\omega_1, \dots, \omega_r]$, since ξ_i^* is integrally dependent on $K[\omega_i]$. Moreover, $\omega_i \equiv 0(\mathfrak{P})$, since $f_i(\xi_i^*) \equiv f_i(v_i^*)(\mathfrak{P})$ and $f_i(v_i^*) = 0$. Let $v_{i1}^* = v_i^*$, $v_{i2}^*, \dots, v_{i, g_i}^*$ be the conjugates of v_i^* over K . We have: $\omega_i = f_i(\xi_i^*) = \prod_{j=1}^{g_i} (\xi_i^* - v_{ij}^*)$. Since $\xi_i^* - v_{ij}^* \not\equiv 0(\mathfrak{P}^*_1)$, if $j \neq 1$, the product $\prod_{j=2}^{g_i} (\xi_i^* - v_{ij}^*)$ is a unit in the quotient ring \mathfrak{Z}^*_1 . Hence

$$\mathfrak{Z}^*_1 \cdot (\omega_1, \dots, \omega_r) = \mathfrak{Z}^*_1 \cdot (\xi_1^* - v_{11}^*, \dots, \xi_r^* - v_{r1}^*) = \mathfrak{P}^*_1,$$

since $\xi_1^* - v_{11}^*, \dots, \xi_r^* - v_{r1}^*$ are uniformizing parameters at P^*_1 . It follows that also $\omega_1, \dots, \omega_r$ are uniformizing parameters at P^*_1 , and consequently they are uniformizing parameters also at P (Theorem 7). We have thus proved the following

THEOREM 8. If P is a simple point, uniformizing parameters $\omega_1, \dots, \omega_r$ at P can be found in such a fashion as to satisfy the conditions: (a) $\omega_i \in \mathfrak{o}$; (b) \mathfrak{o} is integrally dependent on $K[\omega_1, \dots, \omega_r]$.

This is an extension of Theorem 5, except for that part of Theorem 5 which asserts that the uniformizing parameters may be chosen as linear forms in the ξ_i . This part of the theorem is not valid, of course, in the general case.

14. In this and in the following sections we wish to prove the following important theorem:

THEOREM 9. The quotient ring $\mathfrak{Z}(=\mathfrak{o}_{\mathfrak{p}})$ of a simple point P contains the relative algebraic closure of K in Σ .

We shall need several lemmas. Let K' denote, as usual, the relative algebraic closure of K in Σ .

LEMMA 4. K' is contained in the residue class field $K_{\mathfrak{p}}$.

Proof. By the assertion $K' \subseteq K_{\mathfrak{p}}$ we mean the following. We know that the residue class field $K^*_{\mathfrak{p}^*_i}$ at each point P^*_i ($i = 1, 2, \dots, h$) coincides with the ground field K^* . Since P^*_i is a simple point, it follows (Theorem 6) that K^* is algebraically closed in Σ^* , whence $K' \subseteq K^*$. In the homomorphic mapping of $\mathfrak{Z}^*_i(=\mathfrak{o}^*_{\mathfrak{p}^*_i})$ upon $K^*(=\mathfrak{o}^*/\mathfrak{p}^*_i)$, the elements of \mathfrak{Z} are mapped

upon a set of elements which form a subfield $K_p^{(i)}$ of K^* , simply isomorphic to K_p . The assertion of the lemma is to the effect that $K' \subseteq K_p^{(i)}$, for $i = 1, 2, \dots, h$.

The proof is similar to the second part of the proof of Theorem 6. Let $\theta \in K'$, $\theta = \alpha/\beta$, $\alpha, \beta \in \mathfrak{F}$. Let $\beta = \psi_p(\eta_1, \dots, \eta_r) + \dots$, $\psi_p \neq 0$, be the expansion of β at the point P^*_i , in terms of the uniformizing parameters η_1, \dots, η_r at P . We have $\beta \equiv 0(\mathfrak{P}_i^{\rho})$, whence, by (23), $\beta \equiv 0(\mathfrak{P}^{\rho})$. We therefore can write β in the form: $\beta = \sum_{(i)} A_{i_1} \dots i_r \eta_1^{i_1} \dots \eta_r^{i_r}$, $i_1 + \dots + i_r = \rho$, $A_{(i)} \in \mathfrak{F}$. The coefficients $c_{i_1} \dots i_r$ of the form $\psi_p(\eta_1, \dots, \eta_r)$ are obviously the K^* -residues of the elements $A_{(i)} \bmod \mathfrak{P}_i^*$. Since the $A_{(i)}$ are elements of \mathfrak{F} , we conclude that the $c_{(i)}$ belong to $K_p^{(i)}$. For the element α we will have the expansion: $\alpha = \theta \psi_p + \dots$, and by the same argument we deduce that the coefficients $\theta c_{(i)}$, belonging to $K_p^{(i)}$. Since not all the coefficients $c_{(i)}$ are zero, it follows that θ is an element of $K_p^{(i)}$, as was asserted.

LEMMA 5. If $\beta \in \mathfrak{F}$ and if η_1, \dots, η_r are uniformizing parameters at the simple point P , then the power series expansion

$$(27) \quad \beta = \psi_0 + \psi_1(\eta_1, \dots, \eta_r) + \dots$$

of β at P^*_i has all its coefficients in $K_p^{(i)}$.

*Proof.*²⁵ That ψ_0 is an element of $K_p^{(i)}$ is trivial, since $\beta \equiv \psi_0(\mathfrak{P}_i^*)$ and $\beta \in \mathfrak{F}$. We therefore use induction. We assume namely, for every element β in \mathfrak{F} , that the coefficients of $\psi_0, \psi_1, \dots, \psi_{m-1}$ are in $K_p^{(i)}$, and we prove that also the coefficients of ψ_m are in $K_p^{(i)}$.

Let $\phi_\sigma = \sum_{(j)} c_{j_1 \dots j_r} \eta_1^{j_1} \dots \eta_r^{j_r}$, $j_1 + \dots + j_r = \sigma$, be an arbitrary form of degree σ in η_1, \dots, η_r , whose coefficients $c_{(j)}$ are in $K_p^{(i)}$. Let $A_{j_1 \dots j_r}$ be an element of \mathfrak{F} such that $A_{j_1 \dots j_r} \equiv c_{j_1 \dots j_r}(\mathfrak{P}_i^*)$. If we put $\alpha = \sum_{(j)} A_{j_1 \dots j_r} \eta_1^{j_1} \dots \eta_r^{j_r}$, then the expansion of α at P^*_i is of the form:

$$\alpha = \phi_\sigma + \text{terms of higher degree.}$$

Let $\alpha = \phi_\sigma + \phi_{\sigma+1} + \dots$. The form $\phi_{\sigma+\nu}$ depends in an obvious manner on the terms of degree ν of the expansion of the various elements $A_{(j)}$. By our induction, the coefficients of these terms are in $K_p^{(i)}$, if $\nu \leq m-1$. Hence the coefficients of $\phi_{\sigma+1}, \dots, \phi_{\sigma+m-1}$ belong to $K_p^{(i)}$. By the same argument we can find successively elements $\alpha_1, \dots, \alpha_{m-2}$ in \mathfrak{F} such that:

$$\alpha_j = \phi_{\sigma+j}^{(j)} + \phi_{\sigma+j+1}^{(j)} + \dots,$$

where the coefficients of the forms $\phi_{\sigma+j}^{(j)}, \dots, \phi_{\sigma+j+m-1}^{(j)}$ are in $K_p^{(i)}$ and

$$\phi_{\sigma+j} + \phi_{\sigma+j}^{(1)} + \dots + \phi_{\sigma+j}^{(j)} = 0, \quad (j = 1, 2, \dots, m-2).$$

²⁵ We point out that in the course of the proof of the preceding Lemma we have incidentally established the truth of Lemma 5 for the coefficients of the terms of lowest degree of the expansion of β .

If we put $\gamma = \alpha + \alpha_1 + \dots + \alpha_{m-2}$, then the expansion of γ is of the form: $\gamma = \phi_\sigma + g_{\sigma+m-1} +$ terms of higher degree, where $g_{\sigma+m-1}$ is a form of degree $\sigma + m - 1$ in η_1, \dots, η_r with coefficients in $K_p^{(i)}$. We now take successively for ϕ_σ the forms $\psi_1, \psi_2, \dots, \psi_{m-1}$ of (27). We get then elements $\gamma_1, \gamma_2, \dots, \gamma_{m-1}$ such that $\gamma_1 = \psi_1 + g_m +$ terms of degree $> m$, $\gamma_j = \psi_j +$ terms of degree $> m$, $j = 2, \dots, m-1$. Here the coefficients of g_m are elements of $K_p^{(i)}$. Let $\psi_0 = \theta_1 \subset K_p^{(i)}$ and let

$$\omega = \beta - (\gamma_1 + \gamma_2 + \dots + \gamma_{m-1}).$$

The element ω has the following expansion at P^*_i :

$$(28) \quad \omega = \theta_1 + (\psi_m - g_m) + \text{terms of degree } > m.$$

Let $f(\theta_1) = 0$ be the irreducible equation, of degree g , which θ_1 satisfies over K , and let $\theta_2, \dots, \theta_g$ be the conjugates of K . We have $f(\omega) = (\omega - \theta_1)(\omega - \theta_2) \dots (\omega - \theta_g)$, and from (28) it follows immediately that:

$$f(\omega) \equiv (\psi_m - g_m)f'(\theta_1) \pmod{\mathfrak{P}_i^{*m+1}}.$$

Now $f(\omega)$ is an element of \mathfrak{S} and $(\psi_m - g_m)f'(\theta_1)$ is the set of terms of lowest degree in the expansion of $f(\omega)$ at P^*_i . Hence²⁵ the coefficients of the form $(\psi_m - g_m)f'(\theta_1)$ are in $K_p^{(i)}$. Since $\theta_1 \subset K_p^{(i)}$, also $f'(\theta_1)$ belongs to $K_p^{(i)}$, and since the coefficients of g_m are in $K_p^{(i)}$, it follows that the coefficients of ψ_m belong to $K_p^{(i)}$, as was asserted.

15. Our next lemma concerns arbitrary integral domains in which every ideal possesses a finite basis (Hilbert's basis theorem). Let \mathfrak{S} be such an integral domain and let \mathfrak{p} be a maximal prime ideal in \mathfrak{S} . We consider an arbitrary ideal \mathfrak{A} in \mathfrak{S} and its decomposition into maximal primary components. Let q_1, q_2, \dots, q_m be the primary components of \mathfrak{A} whose prime ideals $\mathfrak{p}_1, \mathfrak{p}_2, \dots, \mathfrak{p}_m$ are multiples of \mathfrak{p} . Let q'_1, q'_2, \dots be the remaining primary components of \mathfrak{A} ; $\mathfrak{p}'_1, \mathfrak{p}'_2, \dots$, — their prime ideals. Thus we have:

$$\begin{aligned} \mathfrak{A} &= [q_1, q_2, \dots, q_m; q'_1, q'_2, \dots] \\ \mathfrak{p}_i &\equiv 0(\mathfrak{p}), & (i = 1, 2, \dots, m); \\ \mathfrak{p}'_j &\not\equiv 0(\mathfrak{p}), & (j = 1, 2, \dots). \end{aligned}$$

Let q denote a primary ideal belonging to \mathfrak{p} and let $\Delta_q(\mathfrak{A}, q)$ be the intersection of all the ideals (\mathfrak{A}, q) as q runs through the totality of all primary ideals belonging to \mathfrak{p} .

LEMMA 6. $\Delta_q(\mathfrak{A}, q) = [q_1, q_2, \dots, q_m]$.

Proof. If we assume that the lemma is true for primary ideals \mathfrak{A} , then the lemma follows in general. Namely, we have: $\Delta_q(\mathfrak{A}, q) \subseteq \Delta_q(q_i, q)$, and hence, by your assumption:

$$(29) \quad \Delta_q(\mathfrak{M}, q) \subseteq [q_1, q_2, \dots, q_m].$$

We put $\mathfrak{M}_1 = [q_1, \dots, q_m]$, $\mathfrak{M}'_1 = [q'_1, q'_2, \dots]$. Since \mathfrak{p} is maximal and $\mathfrak{M}'_1 \not\equiv 0(\mathfrak{p})$, we have $(\mathfrak{M}'_1, q) = \mathfrak{S}$, for any primary ideal q belonging to \mathfrak{p} . Consequently, we can write: $\mathfrak{M}_1 = (\mathfrak{M}_1 \mathfrak{M}'_1, \mathfrak{M}_1 q) \equiv 0(\mathfrak{M}, q)$, whence:

$$(29') \quad [q_1, \dots, q_m] = \mathfrak{M}_1 \subseteq \Delta_q(\mathfrak{M}, q).$$

From (29) and (29') our lemma follows.

Let now \mathfrak{M} be a primary ideal, and let \mathfrak{P} be the associated prime ideal. We may assume $\mathfrak{P} \equiv 0(\mathfrak{p})$, because in the contrary case the lemma is trivial. We denote by δ our ideal $\Delta_q(\mathfrak{M}, q)$ and we first establish the following relation:

$$(30) \quad (\delta \mathfrak{p}, \mathfrak{M}) = \delta.$$

Since $\mathfrak{M} \equiv 0(\delta)$, the relation

$$(31) \quad (\delta \mathfrak{p}, \mathfrak{M}) \equiv 0(\delta)$$

is trivial. Let

$$(32) \quad (\delta \mathfrak{p}, \mathfrak{M}) = [q, q'_1, q'_2, \dots]$$

be the decomposition of the ideal $(\delta \mathfrak{p}, \mathfrak{M})$ into maximal primary components, where we assume that the prime ideal \mathfrak{p}'_i associated with q'_i is $\neq \mathfrak{p}$, $i = 1, 2, \dots$, and that q is either the unit ideal or belongs to \mathfrak{p} . Since $\mathfrak{M} \equiv 0(q)$, we have, by definition of δ , that $\delta \equiv 0(q)$. On the other hand $\delta \mathfrak{p} \equiv 0(q'_i)$, and since $\mathfrak{p} \not\equiv 0(\mathfrak{p}'_i)$ it follows that $\delta \equiv 0(q'_i)$. Hence $\delta \subseteq (\delta \mathfrak{p}, \mathfrak{M})$, and (30) follows in view of (31).

By means of the relation (30) the proof of the lemma is readily completed. Let d_1, \dots, d_ρ be a basis for the ideal δ . By (30) we have the following set of relations: $d_i = \sum_{j=1}^{\rho} p_{ij} d_j + a_i$, $i = 1, 2, \dots, \rho$, where the p_{ij} are in \mathfrak{p} and a_1, \dots, a_ρ are elements of \mathfrak{M} . Hence

$$(33) \quad \sum_{j=1}^{\rho} (\delta_{ij} - p_{ij}) d_j \equiv 0(\mathfrak{M}), \quad (i = 1, 2, \dots, \rho),$$

where $\delta_{ij} = 0$ or 1 , according as $i \neq j$ or $i = j$. The determinant $\lambda = |\delta_{ij} - p_{ij}|$ is of the form $1 + p$, $p \equiv 0(\mathfrak{p})$. Hence $\lambda \not\equiv 0(\mathfrak{p})$, whence a fortiori $\lambda \not\equiv 0(\mathfrak{P})$. Since \mathfrak{M} is primary, with \mathfrak{P} as associated prime ideal, we conclude from (33) that d_1, \dots, d_ρ belong to \mathfrak{M} . Hence $\delta \equiv 0(\mathfrak{M})$, and since $\mathfrak{M} \equiv 0(\delta)$, it follows that $\Delta_q(\mathfrak{M}, q) = \delta = \mathfrak{M}$, q. e. d.

16. Now at last we are in position to prove Theorem 9. Let θ be an element of K' . Since \mathfrak{Z} is the quotient field of \mathfrak{F} , we can write $\theta = \alpha/\beta$, $\alpha, \beta \in \mathfrak{F}$. We may assume $\beta \equiv 0(\mathfrak{P})$, because otherwise there is nothing to prove. Consider one of the points P^*_1, \dots, P^*_h , say P^*_1 . By Lemma 4 there exists an element γ_1 in \mathfrak{F} such that $\gamma_1 \equiv \theta(\mathfrak{P}^*_1)$. Let $\gamma_1 = \theta + \psi_1^{(1)} + \psi_2^{(1)} + \dots$

be the expansion of γ_1 at $P^*_{\mathfrak{p}_1}$, in terms of the uniformizing parameters η_1, \dots, η_r at P . Here $\psi_i^{(1)}$ is a form of degree i , and its coefficients are in $K_{\mathfrak{p}}^{(1)}$ (Lemma 5). In particular, the coefficients of the linear form $\psi_1^{(1)}$ are in $K_{\mathfrak{p}}^{(1)}$, and therefore we can find an element γ'_1 in \mathfrak{Z} whose expansion is of the form: $\gamma'_1 = -\psi_1^{(1)} + \dots$. If we put $\gamma_2 = \gamma_1 + \gamma'_1$, then $\gamma_2 = \theta + \psi_2^{(2)} + \text{terms of degree } > 2$, where $\psi_2^{(2)}$ is a quadratic form in η_1, \dots, η_r with coefficients in $K_{\mathfrak{p}}^{(1)}$. Let γ'_2 be an element of \mathfrak{Z} whose expansion is of the form: $\gamma'_2 = -\psi_2^{(2)} + \text{terms of degree } > 2$, and let $\gamma_3 = \gamma_2 + \gamma'_2$. Then γ_3 is an element of \mathfrak{Z} whose expansion at $P^*_{\mathfrak{p}_1}$ is of the form: $\gamma_3 = \theta + \psi_3^{(3)} + \text{terms of degree } > 3$, and again all the coefficients in this expansion belong to $K_{\mathfrak{p}}^{(1)}$ (Lemma 5). Continuing in this manner, we can find for each positive integer i an element γ_i in \mathfrak{Z} whose expansion is of the form: $\gamma_i = \theta + \psi_i^{(i)} + \text{terms of degree } > i$. Here $\psi_i^{(i)}$ is a form of degree i in η_1, \dots, η_r with coefficients in $K_{\mathfrak{p}}^{(1)}$. Hence $\gamma_i \equiv \theta (\mathfrak{P}^{*1}_1)^i$, and since $\alpha = \theta\beta$, it follows that $\alpha - \gamma_i\beta \equiv 0 (\mathfrak{P}^{*1}_1)^i$. Now $\alpha - \gamma_i\beta$ is an element of \mathfrak{Z} and we have proved earlier that $\mathfrak{P}^{*1}_1 \wedge \mathfrak{Z} = \mathfrak{P}^1$ (congruence (23), Section 12). Hence

$$(34) \quad \alpha - \gamma_i\beta \equiv 0 (\mathfrak{P}^i), \quad (i = 1, 2, \dots).$$

Let \mathfrak{Q} be an arbitrary primary ideal belonging to \mathfrak{P} and let ρ be the exponent of \mathfrak{Q} . From (34), for $i = \rho$, we deduce $\alpha \equiv 0 (\beta, \mathfrak{Q})$, whence

$$(35) \quad \alpha \subseteq \underset{\mathfrak{Q}}{\Delta}(\beta, \mathfrak{Q}).$$

Since \mathfrak{P} is maximal, we may apply Lemma 6, where we put $\mathfrak{A} = \mathfrak{Z} \cdot \beta$, $\mathfrak{p} = \mathfrak{P}$. In applying this lemma we must take into account that every ideal in \mathfrak{Z} is a multiple of \mathfrak{P} , whence the primary components q'_1, q'_2, \dots of \mathfrak{A} are never present in the case under consideration. In other words: in the present case our lemma asserts that the intersection of the ideals (β, \mathfrak{Q}) is the ideal $\mathfrak{Z} \cdot \beta$ itself. Hence, in view of (35), we conclude that $\alpha \subset \mathfrak{Z} \cdot \beta$, whence α/β , i. e. θ , is an element of \mathfrak{Z} . This completes the proof of Theorem 9.

The following theorem is a generalization of Theorem 6:

THEOREM 10. *In order that P be a simple point of V_r it is necessary and sufficient that: (1) $P^*_{\mathfrak{p}_1}, \dots, P^*_{\mathfrak{p}_h}$ be simple points of $V^*_{\mathfrak{p}_r}$ and (2) that the quotient ring $\mathfrak{Z}(=\mathfrak{o}_{\mathfrak{p}})$ of P contain the relative algebraic closure K' of K in Σ .*

Proof. We have already proved that the conditions are necessary (see Section 12 and Theorem 9). We prove that they are sufficient. The uniformizing parameters at the simple point $P^*_{\mathfrak{p}_1}$ may be chosen in \mathfrak{Z} (Section 13). Let η_1, \dots, η_r be such uniformizing parameters. We have

$$(36) \quad \mathfrak{Z}^*_{\mathfrak{p}_1}(\eta_1, \dots, \eta_r) = \mathfrak{P}^*_{\mathfrak{p}_1},$$

where $\mathfrak{Z}^*_1 = \mathfrak{o}^*_{\mathfrak{p}^*}$. We now use condition (2). Since $K' \subset \mathfrak{Z}$, we can apply Theorem 2' to the ring \mathfrak{Z} (put $\mathfrak{o} = \mathfrak{Z}$, $\Delta = \Delta' = K'$). Let $\mathfrak{Z}^* = K^*\mathfrak{Z}$ and let $\bar{\mathfrak{P}}_2, \dots, \bar{\mathfrak{P}}_\sigma, \sigma \leq h$ (see (21)), be the conjugates of the prime ideal $\bar{\mathfrak{P}}_1$ under the relative automorphisms of \mathfrak{Z}^* over \mathfrak{Z} . The intersection $[\bar{\mathfrak{P}}_1, \dots, \bar{\mathfrak{P}}_\sigma]$ is an invariant ideal and its contracted ideal in \mathfrak{Z} is \mathfrak{P} . Hence, by Theorem 2', $[\bar{\mathfrak{P}}_1, \dots, \bar{\mathfrak{P}}_\sigma] = \mathfrak{Z}^*\mathfrak{P}$, and consequently $\sigma = h$, i.e. *the prime ideals $\bar{\mathfrak{P}}_1, \dots, \bar{\mathfrak{P}}_h$ in \mathfrak{Z}^* which lie over \mathfrak{P} form a complete set of conjugate ideals*. From (22) and (36) it follows that $\bar{\mathfrak{P}}_1$ is a maximal isolated component of the ideal $\mathfrak{Z}^*(\eta_1, \dots, \eta_r)$. Since this last ideal is invariant, it follows that all the ideals $\bar{\mathfrak{P}}_i, i = 1, 2, \dots, h$, are maximal isolated components of $\mathfrak{Z}^*(\eta_1, \dots, \eta_r)$.²⁶ Taking into account the fact that $\bar{\mathfrak{P}}_1, \dots, \bar{\mathfrak{P}}_h$ are the only maximal prime ideals in \mathfrak{Z}^* , it follows that $\mathfrak{Z}^*(\eta_1, \dots, \eta_r) = \bar{\mathfrak{P}}_1 \cdots \bar{\mathfrak{P}}_h = \mathfrak{Z}^*\mathfrak{P}$. Hence, by Theorem 1,

$$\mathfrak{Z} \cdot (\eta_1, \dots, \eta_r) = \mathfrak{P},$$

and this shows that P is a simple point, q. e. d.

17. In Section 10 we have extended to the case $K_{\mathfrak{p}} = K$ the theorem on the different G'_{ω} proved in our paper.¹⁸ We now propose to prove this theorem in the most general case now under consideration.

Let P be a point of V_r, \mathfrak{p} —the corresponding prime \mathfrak{o} -ideal. Let η_1, \dots, η_r be algebraically independent elements of \mathfrak{p} such that \mathfrak{o} is integrally dependent on the ring $K[\eta_1, \dots, \eta_r]$. Given an element ω in \mathfrak{o} we denote by $G(\eta_1, \dots, \eta_r, z)$ the norm of $z - \omega$ with respect to the field $K(\eta_1, \dots, \eta_r)$.

THEOREM 11. *A necessary and sufficient condition that P be a simple point and that η_1, \dots, η_r be uniformizing parameters at P , is that there exist an element ω such that $G'_{\omega}(\eta_1, \dots, \eta_r; \omega) \not\equiv 0(\mathfrak{p})$.*

We first prove that the condition is necessary. If K' is the relative algebraic closure of K in \mathfrak{Z} , then since P is a simple point, $K' \subseteq K_{\mathfrak{p}}$ (Theorem 9, or Lemma 4). Let

$$[K_{\mathfrak{p}}:K'] = m, [K':K] = \mu, [K^*:K'] = g,$$

whence $[K_{\mathfrak{p}}:K] = m\mu$. Since $K' \subset \mathfrak{Z}$, the ideal $\mathfrak{Z}^*\mathfrak{P}$ decomposes into at most m prime ideals $\bar{\mathfrak{P}}_i$, i.e. we have $h \leq m$ (see Section 4, where we should put $K = K' = \Delta$, $\bar{\Delta} = K^*$, whence $\Delta_{\mathfrak{p}} = K_{\mathfrak{p}}$, since $K_{\mathfrak{p}} \subseteq K^*$). On the other

²⁶ Hence η_1, \dots, η_r are also uniformizing parameters at $P^*_{2^*}, \dots, P^*_{h^*}$. Without the condition $K' \subset \mathfrak{Z}$ this is not always true. For instance, in the example given in footnote ¹¹ let P be the point given by the ideal $\mathfrak{o} \cdot (x, y^2 - 2, z)$. The corresponding points $P^*_{1^*}, P^*_{2^*}$ on $V^*_{\mathfrak{p}}$ are given by the prime ideals $\mathfrak{o}^*(x, y - \sqrt{2})$, $\mathfrak{o}^*(x, y + \sqrt{2})$ respectively. The elements $\eta_1 = y^2 - 2$, $\eta_2 = z + 2x (= \sqrt{2} \cdot x(y + \sqrt{2}))$ are uniformizing parameters at $P^*_{1^*}$ but not at $P^*_{2^*}$.

hand, since K' is algebraically closed in Σ , the prime ideals of $\Sigma^*\mathfrak{P}$ form a set of conjugates under the relative automorphisms of Σ^* over Σ . These automorphisms are extensions of the relative automorphisms of K^* over K' . If we now take into account the relation (6), or (6'), of Section 4, we deduce that $h = m$ and that

$$(37) \quad \Sigma^*\mathfrak{P} = [\bar{\mathfrak{P}}_1 \cdots \bar{\mathfrak{P}}_m] = \bar{\mathfrak{P}}_1 \cdots \bar{\mathfrak{P}}_m,$$

$$(37') \quad \mathfrak{o}^*\mathfrak{p} = [\mathfrak{p}^*_{i_1} \cdots \mathfrak{p}^*_{i_m}] = \mathfrak{p}^*_{i_1} \cdots \mathfrak{p}^*_{i_m}.$$

Let $F(\eta_1, \dots, \eta_r; z)$ denote the norm of $z - \omega$ with respect to the field $K^*(\eta_1, \dots, \eta_r)$, where ω is an element of \mathfrak{o}^* . Our theorem is true for each of the simple points $P^*_{i_1}, \dots, P^*_{i_m}$, in view of the fact that at each of these points the residue class field coincides with the ground field K^* . Thus, dealing with the point $P^*_{i_1}$, we may assert that there exists an element ω in \mathfrak{o}^* such that

$$(38) \quad F'_\omega(\eta_1, \dots, \eta_r; \omega) \not\equiv 0(\mathfrak{p}^*_{i_1}).$$

With the aid of the Remark at the end of Section 10 we proceed to make a judicious choice of the element ω . Let $\theta = \theta_1^{(1)}$ be a primitive element of $K_p^{(1)}$ with respect to K^{27} and let $\theta_i^{(j)}$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, \mu$, be the conjugates of θ over K . We choose the notations in such a fashion that $\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_m^{(j)}$ is a complete set of conjugates with respect to the intermediate field K' . Let $f(\theta) = 0$ be the irreducible equation, of degree $m\mu$, which θ satisfies over K . Since $\theta \in K_p^{(1)}$, there exists an element ξ in \mathfrak{o} such that

$$(39) \quad \xi \equiv \theta_1^{(1)}(\mathfrak{p}^*_{i_1}),$$

whence

$$(39') \quad f(\xi) \equiv 0(\mathfrak{p}).$$

Let

$$\mathfrak{o} \cdot (\eta_1, \dots, \eta_r) = [\mathfrak{p}, \mathfrak{q}_1, \dots, \mathfrak{q}_\sigma]$$

be the decomposition of the ideal $\mathfrak{o} \cdot (\eta_1, \dots, \eta_r)$ into primary maximal components (see (9'), Section 8), and let $\mathfrak{p}_1, \dots, \mathfrak{p}_\sigma$ be prime ideals associated with the primary ideals $\mathfrak{q}_1, \dots, \mathfrak{q}_\sigma$ respectively. The ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_\sigma$ are zero-dimensional, since \mathfrak{o} is integrally dependent on the ring $K[\eta_1, \dots, \eta_r]$. We show that there exists an element ω in \mathfrak{o} such that

$$(40) \quad f(\omega) \equiv 0(\mathfrak{p}),$$

$$(40') \quad f(\omega) \not\equiv 0(\mathfrak{p}_i), \quad (i = 1, 2, \dots, \sigma).$$

²⁷ $K_p^{(1)}$ is a subfield of K^* , simply isomorphic to K_p , and is contained in the residue class field of the point $P^*_{i_1}$. This field has been first introduced in the proof of Lemma 4 (Section 14). The fields $K_p^{(1)}, K_p^{(2)}, \dots, K_p^{(m)}$ are conjugate fields over K' , not necessarily distinct.

We put, namely, $\omega = \zeta + c\pi$, where $c \in K$ and π is an element of \mathfrak{p} but not in \mathfrak{p}_i , $i = 1, 2, \dots, \sigma$. Since $\omega \equiv \zeta(\mathfrak{p})$, the congruence $f(\omega) \equiv 0(\mathfrak{p})$ follows from (39'). On the other hand we have:

$$f(\omega) = f(\zeta + c\pi) = f(\zeta) + c \cdot \pi f'(\zeta) + \dots + c^{m\mu} \pi^{m\mu}$$

(we assume that the leading coefficient of $f(\theta)$ is 1). This is a polynomial in c with coefficients not all $\equiv 0(\mathfrak{p}_i)$, $i = 1, 2, \dots, \sigma$, since $\pi^{m\mu} \not\equiv 0(\mathfrak{p}_i)$. Hence we can find the constant c in K so as to satisfy the relation $f(\omega) \not\equiv 0(\mathfrak{p}_i)$, for all $i = 1, 2, \dots, \sigma$. We assert that the element ω of \mathfrak{o} , constructed in this fashion, necessarily satisfies (38). Namely, let $\mathfrak{p}_{i1}^*, \mathfrak{p}_{i2}^*, \dots$ be the prime ideals in \mathfrak{o}^* which lie over \mathfrak{p}_i . The ideals $\mathfrak{p}_1^*, \dots, \mathfrak{p}_m^*, \mathfrak{p}_{ij}^*$ are then the prime ideals associated with the primary components of the ideal $\mathfrak{o}^*(\eta_1, \dots, \eta_r)$. Since $\mathfrak{p}_2^*, \dots, \mathfrak{p}_m^*$ are the conjugates of \mathfrak{p}_1^* , it follows, by (39), that

$$(39'') \quad \omega \equiv \theta_1^{(1)}(\mathfrak{p}_1^*),$$

whence

$$(41) \quad \omega \not\equiv \theta_1^{(1)}(\mathfrak{p}_i^*), \quad (i = 2, \dots, m).$$

Moreover, by (40'), we have $f(\omega) \not\equiv 0(\mathfrak{p}_{ij}^*)$, $j = 1, 2, \dots$, whence, in particular,

$$(41') \quad \omega \not\equiv \theta_1^{(1)}(\mathfrak{p}_{ij}^*), \quad (i = 1, 2, \dots, \sigma; j = 1, 2, \dots).$$

The relations (39''), (41) and (41') imply (38), in view of the remark at the end of Section 10.

Since K' is algebraically closed in Σ , it follows, by an argument repeatedly used before and based on Lemma 1, that the norm of $z - \omega$ with respect to the field $K^*(\eta_1, \dots, \eta_r)$ is the same as the norm of $z - \omega$ with respect to the field $K'(\eta_1, \dots, \eta_r)$. Therefore the coefficients of $F(\eta_1, \dots, \eta_r, z)$ are in K' . If $F_1, \dots, F_{\mu-1}$ denote the conjugate polynomials of F with respect to K and if $G(\eta_1, \dots, \eta_r; z)$ is the norm of $z - \omega$ with respect to $K(\eta_1, \dots, \eta_r)$, then obviously

$$(42) \quad G(\eta_1, \dots, \eta_r; z) = FF_1 \dots F_{\mu-1}.$$

If we put $F(0, \dots, 0; z) = \phi(z)$, then $\phi(\omega) \equiv 0(\mathfrak{p}_1^*)$, since $F(\eta_1, \dots, \eta_r; \omega) = 0$ and $\eta_i \equiv 0(\mathfrak{p}_1^*)$. Consequently $\phi(z)$ is divisible by $z - \theta_1^{(1)}$ (39''). But (38) implies that $\phi(z)$ is not divisible by $(z - \theta_1^{(1)})^2$. Since $F(\eta_1, \dots, \eta_r, z)$ is invariant under the relative automorphism of Σ^* over Σ (its coefficient being in K'), it follows likewise that $\phi(z)$ is divisible by $z - \theta_i^{(1)}$, but is not divisible by $(z - \theta_i^{(1)})^2$, $i = 1, 2, \dots, m$.

We can also show that $\phi(z)$ is not divisible by $z - \theta_i^{(j)}$, for all $j \neq 1$ and $i = 1, 2, \dots, m$. In fact, if say $\phi(z)$ was divisible by $z - \theta_1^{(2)}$, then $\theta_1^{(2)}$ would have to be the value of ω at some prime ideal \mathfrak{p}^* belonging to the

ideal $\mathfrak{o}^* \cdot (\eta_1, \dots, \eta_r)$,²⁸ i. e. $\omega \equiv \theta_1^{(2)}(\mathfrak{p}^*)$. This ideal \mathfrak{p}^* could not be any of the ideals $\mathfrak{p}_{i_1}^*, \dots, \mathfrak{p}_m^*$, since $\omega \equiv \theta_i^{(1)}(\mathfrak{p}_i^*)$, $i = 1, 2, \dots, m$, and $\theta_1^{(2)} \not\equiv \theta_i^{(1)}$. Hence \mathfrak{p}^* would have to be one of the ideals $\mathfrak{p}_{i_1}^*, \mathfrak{p}_{i_2}^*, \dots$, $i = 1, 2, \dots, \sigma$, considered above. This, however, would be in contradiction with (40').

The polynomial $\phi(z)$ therefore factors as follows:

$$\phi(z) = \prod_{i=1}^m (z - \theta_i^{(1)}) \cdot \psi(z),$$

where $\psi(z)$ and $f(z)$ have no common roots. Let $\phi_j(z) (= F_j(0, \dots, 0, z))$, $j = 1, 2, \dots, \mu - 1$, be the conjugates of $\phi(z)$ with respect to \mathbb{K} . We will have likewise:

$$\phi_j(z) = \prod_{i=1}^m (z - \theta_i^{(j)}) \cdot \psi_j(z), \quad (j = 1, 2, \dots, \mu - 1),$$

where again $\psi_j(z)$ and $f(z)$ have no common roots. By (42) we have:

$$G(0, \dots, 0; z) = \phi(z)\phi_1(z) \cdots \phi_{\mu-1}(z) = f(z) \cdot A(z),$$

where $A(z) (= \psi\psi_1 \cdots \psi_{\mu-1})$ and $f(z)$ are relatively prime. Since $\eta_i \equiv 0(\mathfrak{p})$, we have: $G'_\omega \equiv f'(\omega) \cdot A(\omega) + f(\omega) \cdot A'(\omega)(\mathfrak{p})$, i. e. $G'_\omega \equiv f'(\omega) \cdot A(\omega)(\mathfrak{p})$, since $f(\omega) \equiv 0(\mathfrak{p})$ (40). Now we observe that $f'(\omega) \not\equiv 0(\mathfrak{p})$, since $f(\omega) \equiv 0(\mathfrak{p})$ and f is an irreducible polynomial, and we also note that $A(\omega) \not\equiv 0(\mathfrak{p})$, since $A(z)$ is not divisible by $f(z)$. Hence $G'_\omega \not\equiv 0(\mathfrak{p})$, as was asserted.

18. Continuation of the proof. The condition is sufficient. From (42) we deduce in the first place that $G'_\omega \not\equiv 0(\mathfrak{p})$ implies the relations $F'_\omega \not\equiv 0(\mathfrak{p}_i^*)$, $i = 1, 2, \dots, h$. Hence from the hypothesis $G'_\omega \not\equiv 0(\mathfrak{p})$ follows at any rate that the points P_1^*, \dots, P_h^* are simple (by the special case $\mathbb{K}_\mathfrak{p} = \mathbb{K}$; see Section 10). It remains to prove that $\mathbb{K}' \subset \mathfrak{F}$ (Theorem 10). We shall prove the following stronger result: \mathfrak{F} is integrally closed in Σ . Since $\mathbb{K} \subset \mathfrak{F}$ and \mathbb{K}' is an algebraic extension of \mathbb{K} , the property of \mathfrak{F} being integrally closed will obviously imply that \mathbb{K}' is a subfield of \mathfrak{F} . Now to show that \mathfrak{F} is integrally closed in Σ , we consider the complementary module \mathfrak{e} ²⁹ of the ring $\mathbb{K}[\eta_1, \dots, \eta_r, \omega]$. If ν denotes the relative degree $[\Sigma: \mathbb{K}(\eta_1, \dots, \eta_r)]$, then it is well known that the elements

$$1/G'_\omega, \omega/G'_\omega, \dots, \omega^{\nu-1}/G'_\omega$$

form a module basis for \mathfrak{e} with respect to the ring $\mathbb{K}[\eta_1, \dots, \eta_r]$. Since $G'_\omega \not\equiv 0(\mathfrak{p})$, it follows therefore that \mathfrak{e} is contained in \mathfrak{F} . Since \mathfrak{e} contains

²⁸ This assertion has been proved in the case of an algebraically closed ground field (18, p. 263, footnote 12) and therefore is obviously true for any ground field.

²⁹ The module \mathfrak{e} consists of those elements ζ of Σ for which the trace $T(\zeta \cdot \mathfrak{a})$ is in $\mathbb{K}[\eta_1, \dots, \eta_r]$, for every element \mathfrak{a} in $\mathbb{K}[\eta_1, \dots, \eta_r, \omega]$.

the integral closure of the ring $K[\eta_1, \dots, \eta_r]$ and since \mathfrak{o} is integrally dependent on this ring, we conclude that \mathfrak{S} contains the integral closure $\bar{\mathfrak{o}}$ of \mathfrak{o} . Let $\bar{\mathfrak{p}}$ be the contracted ideal of \mathfrak{P} in $\bar{\mathfrak{o}}$: $\bar{\mathfrak{p}} = \mathfrak{P} \cap \bar{\mathfrak{o}}$. We have then $\mathfrak{S} = \mathfrak{S}_{\bar{\mathfrak{p}}} \supseteq \bar{\mathfrak{o}}_{\bar{\mathfrak{p}}}$. On the other hand $\bar{\mathfrak{o}}_{\bar{\mathfrak{p}}} \supseteq \mathfrak{o}_{\bar{\mathfrak{p}}}$ (since $\bar{\mathfrak{p}} \cap \mathfrak{o} = \mathfrak{p}$), whence $\bar{\mathfrak{o}}_{\bar{\mathfrak{p}}} \supseteq \mathfrak{S}$. Consequently $\mathfrak{S} = \bar{\mathfrak{o}}_{\bar{\mathfrak{p}}}$, and therefore \mathfrak{S} is integrally closed (since for the integrally closed ring $\bar{\mathfrak{o}}$ it is true that the quotient ring of any prime $\bar{\mathfrak{o}}$ -ideal is integrally closed). This completes the proof of Theorem 11.

As a corollary we have the following

THEOREM 12. *The quotient ring \mathfrak{S} of a simple point is integrally closed in its quotient field.*

Remark. If P is a simple point, with η_1, \dots, η_r as uniformizing parameters, and if \mathfrak{o} is integrally dependent on $K[\eta_1, \dots, \eta_r]$, then the elements ω of \mathfrak{o} such that $G'_\omega \not\equiv 0(\mathfrak{p})$ are characterized by the relations (40), (40'), where the irreducible polynomial $f(\omega)$ must be of degree $m_\mu = [K_\mathfrak{p} : K]$. This follows from the first part of the proof of Theorem 11 (Section 17) and from the remark at the end of Section 10.

In Theorem 11 we have assumed that the elements η_1, \dots, η_r are in \mathfrak{p} . We now drop this assumption, and we consider the uniquely determined irreducible polynomials $f_i(z)$ in $K[z]$ ($i = 1, 2, \dots, r$) such that $f_i(\eta_i) \equiv 0(\mathfrak{p})$. We can easily prove the following stronger theorem:

THEOREM 11'. *A necessary and sufficient condition that P be a simple point and that $f_1(\eta_1), \dots, f_r(\eta_r)$ be uniformizing parameters at P is that there exist an element ω in \mathfrak{o} such that $G'_\omega(\eta_1, \dots, \eta_r; \omega) \not\equiv 0(\mathfrak{p})$.*

That the condition is necessary follows almost immediately from Theorem 11. Namely, let us put $\xi_i = f_i(\eta_i)$ and let $H(\xi_1, \dots, \xi_r; z)$ denote the norm of $z - \omega$ with respect to the field $K(\xi_1, \dots, \xi_r)$, while $G(\eta_1, \dots, \eta_r; z)$ denotes, as before, the norm of $z - \omega$ with respect to the field $K(\eta_1, \dots, \eta_r)$. Since this last field contains the field $K(\xi_1, \dots, \xi_r)$, it is clear that we have an identity (in z) of the form:

$$H(\xi_1, \dots, \xi_r; z) = G(\eta_1, \dots, \eta_r; z) \cdot A(\eta_1, \dots, \eta_r; z),$$

where A is a polynomial with coefficients in K . Since $G(\eta_1, \dots, \eta_r; \omega) = 0$ we have

$$H'_\omega(\xi_1, \dots, \xi_r; \omega) = G'_\omega(\eta_1, \dots, \eta_r; \omega) \cdot A(\eta_1, \dots, \eta_r; \omega).$$

Now, by hypothesis, ξ_1, \dots, ξ_r are uniformizing parameters at P , and moreover, \mathfrak{o} depends integrally on the ring $K[\xi_1, \dots, \xi_r]$, since each element η_i is integrally dependent on $K[\xi_i]$. Hence, by Theorem 11, we must have $H'_\omega \not\equiv 0(\mathfrak{p})$, for a suitable element ω in \mathfrak{o} . The above identity shows then that we must also have $G'_\omega \not\equiv 0(\mathfrak{p})$, as was asserted.

The proof that the condition is sufficient is direct and follows the lines of the proof of sufficiency given in the special case of Theorem 11. The relation $G'_\omega \not\equiv 0(\mathfrak{p})$ implies the relations $F'_\omega(\eta_1, \dots, \eta_r; \omega) \not\equiv 0(\mathfrak{p}^*_i)$, for $i = 1, 2, \dots, h$. Hence P^*_1, \dots, P^*_h are simple points and $\eta_1 - \theta_1^{(i)}, \dots, \eta_r - \theta_r^{(i)}$ are uniformizing parameters at P^*_i , where the $\theta_i^{(j)}$ are elements of K^* and $\eta_j \equiv \theta_j^{(i)}(\mathfrak{p}^*_i)$ (Section 10). Since $f_j(\eta_j) \equiv 0(\mathfrak{p}) \equiv 0(\mathfrak{p}^*_i)$, $\theta_j^{(i)}$ must be a root of $f_j(z)$. Since $f_j(z)$ has no repeated roots, it follows that $f_j(\eta_j)$ differs from $\eta_j - \theta_j^{(i)}$ by a factor which is a unit in the quotient ring \mathfrak{S}^*_i of the point P^*_i . Hence $f_1(\eta_1), \dots, f_r(\eta_r)$, i. e. ξ_1, \dots, ξ_r are also uniformizing parameters at P^*_i ($i = 1, 2, \dots, h$). It remains to prove that $K' \subset \mathfrak{S}$, since from this it will follow that P is a simple point (Theorem 10) and that ξ_1, \dots, ξ_r are uniformizing parameters at P (Theorem 7). The rest of the proof is the same as before, namely it is shown that \mathfrak{S} is integrally closed in Σ .

COROLLARY. If V_r is a linear space, i. e. if \mathfrak{o} is a polynomial ring, then every point of V_r is simple.

In fact, if $\mathfrak{o} = K[\xi_1, \dots, \xi_r]$, where ξ_1, \dots, ξ_r are algebraically independent elements, then the norm G of $z - \omega$ with respect to the field $K(\xi_1, \dots, \xi_r)$ is $z - \omega$ itself, whence $G'_\omega = 1$.

V. Simple subvarieties.

19. Let V_s be an irreducible *simple* subvariety of V_r , of dimension s . Let \mathfrak{p} be the corresponding s -dimensional prime ideal in \mathfrak{o} and let $\mathfrak{S} = \mathfrak{o}_{\mathfrak{p}}$. By definition, there exist $r - s$ elements in \mathfrak{S} , say $\eta_1, \dots, \eta_{r-s}$, such that

$$(43) \quad \mathfrak{S} \cdot (\eta_1, \dots, \eta_{r-s}) = \mathfrak{P} = \mathfrak{S} \cdot \mathfrak{p}.$$

Let ξ_1, \dots, ξ_s be elements of \mathfrak{S} which are algebraically independent mod \mathfrak{p} (with respect to the ground field K), but otherwise arbitrary. We take as new ground field the field $\Omega = K(\xi_1, \dots, \xi_s)$ and we put $\bar{\mathfrak{o}} = \Omega \cdot \mathfrak{o}$, $\bar{\mathfrak{p}} = \bar{\mathfrak{o}} \cdot \mathfrak{p}$. Since Ω is a pure transcendental extension of K , $\bar{\mathfrak{p}}$ is prime. It is of dimension zero, since the ξ 's are algebraically independent mod \mathfrak{p} . With Ω as new ground field, the elements ξ_1, \dots, ξ_s define an $(r - s)$ -dimensional variety \bar{V}_{r-s} of which they are the coördinates of the general point. The ideal $\bar{\mathfrak{p}}$ defines a point \bar{P} on \bar{V}_{r-s} . The quotient ring $\bar{\mathfrak{S}} = \bar{\mathfrak{o}}_{\bar{\mathfrak{p}}}$ coincides with \mathfrak{S} , since the elements of Ω are units in \mathfrak{S} . Hence, by (43),

$$(43') \quad \bar{\mathfrak{S}} \cdot (\eta_1, \dots, \eta_{r-s}) = \bar{\mathfrak{S}} \cdot \bar{\mathfrak{p}}.$$

i. e. \bar{P} is a simple point of \bar{V}_{r-s} .

Conversely, assume that \bar{P} is a simple point of \bar{V}_{r-s} . There will exist then elements $\eta_1, \dots, \eta_{r-s}$ in $\bar{\mathfrak{S}}$, i. e. in \mathfrak{S} , satisfying (43'). The relation (43)

is equivalent to (43'). Hence V_s is a simple subvariety of V_r . Thus the assertions: V_s is a simple subvariety of V_r ; \tilde{P} is a simple point of \tilde{V}_{r-s} —are equivalent.

THEOREM 13. *The quotient ring $\mathfrak{S}(=\mathfrak{o}_{\tilde{P}})$ of a simple subvariety is integrally closed in Σ .*

The theorem is an immediate consequence of Theorem 12 and of the fact that $\mathfrak{S} = \tilde{\mathfrak{S}} = \bar{\mathfrak{o}}_{\tilde{P}}$.

THEOREM 14. *If V_s is a simple subvariety of V_r , the uniformizing parameters $\eta_1, \dots, \eta_{r-s}$ along V_s and the elements ξ_1, \dots, ξ_s algebraically independent mod \mathfrak{p} , can be so chosen that they be elements of \mathfrak{o} and that \mathfrak{o} be integrally dependent on the ring $K[\eta_1, \dots, \eta_{r-s}; \xi_1, \dots, \xi_s]$.*

Proof. Let $\bar{\mathfrak{o}} = \mathfrak{o}/\mathfrak{p} = K[\bar{\xi}_1, \dots, \bar{\xi}_s]$, whence $\bar{\mathfrak{o}}$ is of degree of transcendency s over K . Subject to a preliminary linear homogeneous transformation on ξ_1, \dots, ξ_s , with coefficients u_{ij} in K , we may assume that the following conditions are satisfied.

- (a) ξ_{r+1}, \dots, ξ_n are integrally dependent on $K[\xi_1, \dots, \xi_r]$;
- (b) $\bar{\xi}_{s+1}, \dots, \bar{\xi}_n$ are integrally dependent on $K[\bar{\xi}_1, \dots, \bar{\xi}_s]$;
- (c) $f_i(\xi_1, \dots, \xi_s; \xi_{s+i})$, $i = 1, 2, \dots, r-s$, are uniformizing parameters along V_s , where f_i is an irreducible polynomial in $\xi_1, \dots, \xi_s, \xi_{s+i}$ with coefficients in K .

The possibility of satisfying conditions (a) and (b) is trivial. As to condition (c), we observe that by (b) the elements ξ_1, \dots, ξ_s are algebraically independent mod \mathfrak{p} . We put $\zeta_i = \xi_i$, $i = 1, 2, \dots, s$. It then follows from the proof of Theorem 8 (Section 13) that for "non special" u_{ij} , certain polynomials $f_i(\xi_{s+i})$, $i = 1, 2, \dots, r-s$, with coefficients in $\Omega(=K(\zeta_1, \dots, \zeta_s))$ will be uniformizing parameters at the point \tilde{P} of \tilde{V}_{r-s} , hence also along V_s . The values of the u_{ij} to be avoided are those which satisfy certain algebraic relations with coefficients in Ω . Since K contains infinitely many elements, the u_{ij} may be chosen in K . We may assume that the leading coefficient of f_i is 1. If we put $\eta_i = f_i(\xi_1, \dots, \xi_s; \xi_{s+i})$, $i = 1, 2, \dots, r-s$, then ξ_1, \dots, ξ_s ($= \xi_1, \dots, \xi_s$) are algebraically independent mod \mathfrak{p} and $\eta_1, \dots, \eta_{r-s}$ are uniformizing parameters along V_s . Since $\eta_i \equiv 0(\mathfrak{p})$, we find, passing to the ring $\bar{\mathfrak{o}}$: $f_i(\bar{\xi}_1, \dots, \bar{\xi}_s; \bar{\xi}_{s+i}) = 0$. Since f_i is irreducible, it follows by condition (b) that the coefficients of f_i are polynomials in ξ_1, \dots, ξ_s . Hence ξ_{s+i} is integrally dependent on $K[\xi_1, \dots, \xi_s; \eta_i]$, $i = 1, 2, \dots, r-s$, whence

ξ_1, \dots, ξ_r are integrally dependent on the ring $K[\xi_1, \dots, \xi_s; \eta_1, \dots, \eta_{r-s}]$. This completes the proof of the theorem, in view of condition (a).

20. Let η_1, \dots, η_r be algebraically independent elements of \bar{o} such that o is integrally dependent on the ring $K[\eta_1, \dots, \eta_r]$. The residual class ring $\bar{o} = o/p$ will depend integrally on the ring $K[\tilde{\eta}_1, \dots, \tilde{\eta}_r]$, and since \bar{o} is of degree of transcendency s over K , s elements $\tilde{\eta}_i$ have to be algebraically independent. Consequently s of the elements η_i are algebraically independent mod p . We assume that η_1, \dots, η_s are algebraically independent mod p . Let $f_i(\eta_1, \dots, \eta_s; \eta_{s+i}) \equiv 0(p)$ be the irreducible congruence which η_{s+i} satisfies over $K[\eta_1, \dots, \eta_s]$. Let moreover, $F(\eta_1, \dots, \eta_r; z)$ be the norm of $z - \omega$ ($\omega \in o$), over the field $K(\eta_1, \dots, \eta_r)$. As a generalization of Theorem 11', we prove the following

THEOREM 15. *The existence of an element ω in o such that $F'_\omega(\eta_1, \dots, \eta_r; \omega) \not\equiv 0(p)$ is a necessary and sufficient condition in order that V_s be a simple subvariety of V_r and that $f_1(\eta_1, \dots, \eta_s, \eta_{s+1}), \dots, f_{r-s}(\eta_1, \dots, \eta_s; \eta_r)$ be uniformizing parameters along V_s .*

The theorem is an immediate consequence of Theorem 11'. It is sufficient to observe that $F(\eta_1, \dots, \eta_r; z)$ is also the norm of $z - \omega$ with respect to the field $\Omega(\eta_{s+1}, \dots, \eta_r)$, where $\Omega = K(\eta_1, \dots, \eta_s)$. Moreover, as was pointed out in the preceding section, the subvariety V_s of V_r and the point \bar{P} of \bar{V}_{r-s} (the elements η_1, \dots, η_s now play the rôle of ξ_1, \dots, ξ_s) are both simple or not simple at the same time, and that uniformizing parameters along V_s are also uniformizing parameters at \bar{P} , and conversely (see (43) and (43')).

An immediate corollary of Theorem 15 is the following: if V_s contains a simple point P of V_r , then V_s itself is a simple subvariety. In fact, if p_0 is the prime zero-dimensional o -ideal which corresponds to the point P , then $p \equiv 0(p_0)$, and $G'_\omega \not\equiv 0(p_0)$ (Theorem 11') implies the relation $G'_\omega \not\equiv 0(p)$.

We can invert this result. We show namely that a simple subvariety V_s contains at least one simple point of V_r . Using the notations of Theorem 15, if V_s is simple, then $G'_\omega \not\equiv 0(p)$, for some ω in o . We can therefore find a prime zero-dimensional divisor p_0 of p such that $G'_\omega \not\equiv 0(p_0)$.³⁰ If P is the point of V_r defined by p_0 , then P is simple (Theorem 11') and lies on V_s (since $p \equiv 0(p_0)$).

THE JOHNS HOPKINS UNIVERSITY.

³⁰ If we pass to the ring o/p , then our assertion is equivalent to the following: if $\alpha \not\equiv 0$, then there exists a zero-dimensional prime ideal which does not contain α . The proof of this assertion is straightforward.

ASSOCIATIVE MULTIPLICATIVE SYSTEMS.*

By J. E. EATON.

1. Introduction. Grouplike systems with non-unique multiplication were first studied by Marty in 1934.¹ Others who have been interested in such systems are Wall, Kuntzmann, Ore, Griffiths, and Krasner.² In 1938 Dresher and Ore³ undertook an axiomatic investigation of the general properties of such systems which they called multigroups.

Many of the results of Dresher and Ore were concerned with the relation of subsets of the multigroup to the multigroup itself. In this paper we shall extend some of these results to algebraic systems with a multivalued associative operation multiplication which however need satisfy no quotient law. As in multigroups, coset decompositions of the system are possible, and we may characterize completely the homomorphisms generated by such decompositions. Normal subsets of the system exist which have interesting structure properties. In particular for these subsets we may enunciate a Jordan-Hölder theorem. The theorems derived in this paper in a few instances represent improvements in the known theorems of the theory of multigroups.

2. Definitions. An *associative multiplicative system* is an algebraic system in which there is defined a single binary operation multiplication. We shall for brevity throughout this paper refer to such a system as an *m-system*. We shall denote by \mathfrak{M} an arbitrary *m-system* and by m_1, m_2, \dots the elements

* Received April 26, 1939.

¹ F. Marty, (1) "Sur une généralisation de la notion de groupe," *Huitième congrès des mathématiciens scandinaves*, Stockholm, 1934, pp. 45-49; (2) "Rôle de la notion d'hypergroupe dans l'étude des groupes non abéliens," *Comptes rendus*, vol. 201 (Paris, 1935), pp. 636-638; (3) "Sur les groupes et hypergroupes attachés à une fraction rationnelle," *Annales de l'école normale*, 3 sér., vol. 53 (1936), pp. 82-123.

² H. S. Wall, "Hypergroups," *American Journal of Mathematics*, vol. 59 (1937), pp. 77-98; J. Kuntzmann, (1) "Opérations multiformes. Hypergroupes," *Comptes rendus*, vol. 204 (Paris, 1937), pp. 1787-1788; (2) "Homomorphie entre systèmes multiformes," *Comptes rendus*, vol. 205 (Paris, 1937), pp. 208-210; (3) "Systèmes multiformes et systèmes hypercomplexes," *Comptes rendus*, vol. 208 (Paris, 1939), pp. 493-495; Oystein Ore, "Structures and group theory, I," *Duke Mathematical Journal*, vol. 3 (1937), pp. 149-174; L. W. Griffiths, "On hypergroups, multigroups, and product systems," *American Journal of Mathematics*, vol. 60 (1938), pp. 345-354; M. Krasner, "Sur la primitivité des corps \mathfrak{P} -adiques," *Mathematica*, vol. 13 (1937), pp. 72-191.

³ Dresher and Ore, "Theory of multigroups," *American Journal of Mathematics*, vol. 60 (1938), pp. 705-733. We shall in the following cite this paper as D. and O.

of \mathfrak{M} . We make no assumption on the finiteness of the number of elements of \mathfrak{M} but we shall suppose that they are at least enumerable. The multiplication which is defined in an m -system is subject to but two axioms: the existence of the product of any two elements and the associative law.

AXIOM 1. *The Product.* If m_i and m_j are any two elements of \mathfrak{M} , then the product $m_i m_j$ is a non-void subset of \mathfrak{M} .

$$m_i m_j = \{m'_k\}.$$

The existence of the product of any two elements of \mathfrak{M} permits us to give meaning to the notion of the product of any two subsets of \mathfrak{M} . If \mathfrak{A} and \mathfrak{B} are two non-void subsets of \mathfrak{M} with elements $\{a_i\}$ and $\{b_i\}$ respectively, then an element m of \mathfrak{M} is in $\mathfrak{A}\mathfrak{B}$ if and only if m is contained in some product $a_j b_k$. The relation of containing we shall symbolize in the usual manner. If \mathfrak{A} and \mathfrak{B} are any two subsets of \mathfrak{M} , $\mathfrak{A} \supset \mathfrak{B}$ shall mean that every element of \mathfrak{B} is an element of \mathfrak{A} .

AXIOM 2. *The Associative Law.* If m_i, m_j, m_k are any three elements of \mathfrak{M} , then

$$(m_i m_j) m_k = m_i (m_j m_k) = m_i m_j m_k.$$

These products have meaning according to the definition of products of subsets.

Kuntzmann ⁴ has noted several weaker forms of this associative law which are of interest in systems in which a non-unique multiplication is defined. We shall have occasion to refer later to one of these.

AXIOM 2'. *The Left Associative Law.* If m_i, m_j, m_k are any three elements of \mathfrak{M} , then

$$(m_i m_j) m_k \supset m_i (m_j m_k).$$

A multigroup is defined by Dresher and Ore ⁵ to be an m -system satisfying the following quotient law.

AXIOM 3. *The Quotient Axiom.* For any ordered pair of elements m_i, m_j there exist at least two elements x and y such that

$$x m_i \supset m_j \quad m_i y \supset m_j.$$

3. Representation. Before we develop some of the properties of an m -system it would be well to derive a representation of such a system. Let \mathfrak{M} be some multiplicative system satisfying Axiom 1 (we do not assume it is associative) and consisting of elements $\{m_i\}$. To each m_i associate a matrix M_i defined in the following manner. If $m_i m_j \supset m_k$, then M_i has e , the unit of a Boolean ring, as its k, j -th element. The remaining elements of M_i are

⁴ *Op. cit.*, 1, p. 1787.

⁵ D. and O., pp. 706-707.

zeros. We call the set of M_i 's the *left regular matrix association*⁶ of \mathfrak{M} . Similarly \bar{M}_i is an element of the right regular matrix association of \mathfrak{M} if when $m_j m_i \supset m_k$ then \bar{M}_i has e in its k, j -th place. We say that a matrix association is a *representation* if when $m_i m_j = \{m'_k\}$, then $M_i M_j = \Sigma M'_k$, where multiplication and addition of the matrices is defined in the usual manner. We then have the following theorem.

THEOREM 1. *The necessary and sufficient condition that a multiplicative system \mathfrak{M} be associative is that both its left and right regular matrix associations are representations.*

Proof. (i) Necessary. Let a, b be elements of \mathfrak{M} , $ab = \{c_1, c_2, \dots\}$, and A, B, C_i the corresponding matrices of the left regular matrix association. Let there be e in the i, k -th place of AB . Then for some j there is e in the i, j -th place of A and in the j, k -th place of B . Hence there is an m_j such that $am_j \supset m_i$ and $bm_k \supset m_j$. Then $abm_k \supset m_i$. Hence for some c_i , $c_i m_k \supset m_i$. Therefore some C_i has e in its i, k -th place. Conversely let there be e in the i, k -th place of some C . Then $cm_k \supset m_i$. Hence $abm_k \supset m_i$. Then for some $m_j \subset bm_k$, $am_j \supset m_i$. Hence A has e in the i, j -th place and B has e in the j, k -th place. Therefore AB has e in the i, k -th place.

(ii) Sufficient. Let the left regular matrix association of \mathfrak{M} be a representation. Consider $m_i(m_j m_k) \supset m_r$. Then for some $m_s \subset m_j m_k$, $m_i m_s \supset m_r$. Hence M_j has e in the s, k -th place and M_i has e in the r, s -th place. Then $M_i M_j$ has e in the r, k -th place. This implies that for some $m \subset m_i m_j$, $mm_k \supset m_r$. We then have $m_i(m_j m_k) \subset (m_i m_j)m_k$. The reverse inequality is obtained if the right regular matrix association is a representation. Hence we have proved the theorem. However the left regular matrix association being a representation does not imply the right is. This may be shown by the system of two elements, a and b , whose multiplication scheme is: $aa = a$; $ab = b$; $ba = a, b$; $bb = b$.

COROLLARY. *The necessary and sufficient condition that a multiplicative system be left associative (satisfy Axiom 2') is that its left regular matrix association is a representation.*

4. Coset decompositions. We shall say that any subset \mathfrak{A} of an m -system is a *subsystem* if $\mathfrak{A} \supset \mathfrak{A}\mathfrak{A}$. We then have immediately that \mathfrak{A} itself is an m -system. The subsystems with which we shall be primarily concerned in this paper are the left reversible subsystems.⁷ A subsystem \mathfrak{A} of an m -system is *left reversible* if when $\mathfrak{A}m_i \supset m_j$, then $\mathfrak{A}m_j \supset m_i$. Similarly \mathfrak{A} is

⁶ Cf. Wall, *op. cit.*, p. 86; D. and O., p. 709.

⁷ The concept of reversibility was introduced in D. and O., p. 715.

right reversible if when $m_i\mathfrak{A} \supset m_j$, then $m_j\mathfrak{A} \supset m_i$. The set $\mathfrak{A}m_i$ we call a *left coset* of \mathfrak{A} . The interesting feature of left reversible subsystems of an m -system is that their left cosets constitute a partition of the m -system in the precise sense stated in the following theorem.⁸

THEOREM 2. *Let \mathfrak{A} be a left reversible subsystem of an m -system \mathfrak{M} . Then \mathfrak{M} has a unique left coset expansion $\mathfrak{M} = \mathfrak{A}m_1 + \mathfrak{A}m_2 + \cdots$ such that*

- (i) *any element in a coset generates the same coset;*
- (ii) *a coset contains its generating element;*
- (iii) *every element of \mathfrak{M} lies in some coset;*
- (iv) *the cosets are disjoint.*

Proof. (i) Consider any coset $\mathfrak{A}m_i$ and any m_j contained in it. Then there is an a_i in \mathfrak{A} such that $a_im_i \supset m_j$. From left reversibility there is an a_j in \mathfrak{A} such that $a_jm_j \supset m_i$. Since \mathfrak{A} is a subsystem, $\mathfrak{A} \supset \mathfrak{A}a$ for all a in \mathfrak{A} . Hence $\mathfrak{A}m_i \supset \mathfrak{A}a_im_i \supset \mathfrak{A}m_j \supset \mathfrak{A}a_jm_j \supset \mathfrak{A}m_i$. Since the first and last elements of the chain are identical, we must have equality throughout. Thus $\mathfrak{A}m_i = \mathfrak{A}m_j$ for any m_j in $\mathfrak{A}m_i$.

- (ii) m_i is in $\mathfrak{A}m_i$ since m_i is in $\mathfrak{A}m_j$ and $\mathfrak{A}m_i = \mathfrak{A}m_j$.
- (iii) From (ii) every m_i in \mathfrak{M} lies in the coset $\mathfrak{A}m_i$.
- (iv) Let m_k be contained in both $\mathfrak{A}m_i$ and $\mathfrak{A}m_j$. Then $\mathfrak{A}m_i = \mathfrak{A}m_k = \mathfrak{A}m_j$.

COROLLARY. *The theorem is true under the weaker assumption that \mathfrak{M} is a left associative multiplicative system.*

We may mention that \mathfrak{M} has a unique double coset expansion $\mathfrak{M} = \mathfrak{A}m_1\mathfrak{B} + \mathfrak{A}m_2\mathfrak{B} + \cdots$ with respect to a left reversible subsystem \mathfrak{A} and a right reversible subsystem \mathfrak{B} .⁹

5. Homomorphisms. Let us consider any partition of an m -system \mathfrak{M} into subsets X_1, X_2, \cdots , where the subsets are not necessarily disjoint. We do however suppose that every element of \mathfrak{M} lies in some X_i . We have already noted what we mean by the product X_iX_j in the element sense. We may define the set product X_iX_j to be the totality of those subsets X_k which have at least one of their elements in the element product X_iX_j . The X_i 's then obviously form an m -system which is homomorphic to the original m -system. By a *homomorphism* of an m -system \mathfrak{M} to an m -system \mathfrak{M}^* we mean the usual many-one correspondence between the elements of \mathfrak{M} and the elements of \mathfrak{M}^* which preserves multiplication. That is, every element of \mathfrak{M} corresponds to a unique image element of \mathfrak{M}^* and every element of \mathfrak{M}^* is the image of at least one element of \mathfrak{M} . Furthermore if m_i, m_j, m_k have the

⁸ This is a direct analogue for m -systems of Theorems 8 and 9, D. and O., p. 717.

⁹ Cf. D. and O., p. 718.

respective images m^*_i , m^*_j , m^*_k and if $m_i m_j \supset m_k$ then $m^*_i m^*_j \supset m^*_k$. Also if $m^*_i m^*_j \supset m^*_k$ then for some m_i , m_j , m_k corresponding to them $m_i m_j \supset m_k$. However, in the homomorphisms which we shall consider we shall find it convenient to place certain restrictions on the inverse correspondence from \mathfrak{M}^* to \mathfrak{M} . A homomorphism from \mathfrak{M} to \mathfrak{M}^* is a *strong left unit homomorphism* if \mathfrak{M}^* contains left scalar units and if for any m and m' with the same image there is an element a in \mathfrak{M} corresponding to some left scalar unit of \mathfrak{M}^* such that $am \supset m'$.

An element e of an m -system \mathfrak{M} is a *left scalar unit*¹⁰ if $em = m$ for every m in \mathfrak{M} . An element which is both a left scalar unit and a right scalar unit ($me = m$) is called an *absolute unit*. Obviously there can be at most one absolute unit in any m -system. A set of left scalar units of an m -system \mathfrak{M}^* form a system of *fundamental units* with respect to a strong left unit homomorphism of \mathfrak{M} to \mathfrak{M}^* if for any one of them there is an a in \mathfrak{M} corresponding to it such that for some m and m' with the same image $am \supset m'$ and if for any m and m' with the same image there is an a corresponding to one of them such that $am \supset m'$. We are now in a position to characterize completely the homomorphisms generated by the left coset decompositions of left reversible subsystems.¹¹

THEOREM 3. *Let \mathfrak{A} be a left reversible subsystem of an m -system \mathfrak{M} . Then \mathfrak{M} is strongly left unit homomorphic to the m -system of the left coset expansion of \mathfrak{M} with respect to \mathfrak{A} , $\mathfrak{M}/\mathfrak{A}$, and the homomorphism has as a set of fundamental units those cosets containing elements of \mathfrak{A} . Conversely by any strong left unit homomorphism $\mathfrak{M} \rightarrow \mathfrak{M}^*$ those elements of \mathfrak{M} which correspond to any set of fundamental units of \mathfrak{M}^* form a left reversible subsystem \mathfrak{A} of \mathfrak{M} such that the m -system of the left coset expansion $\mathfrak{M}/\mathfrak{A}$ is isomorphic to \mathfrak{M}^* .*

Proof. Since \mathfrak{A} is a subsystem, $\mathfrak{A} \supset \mathfrak{A}a$ for any a in \mathfrak{A} . This implies $\mathfrak{A}\mathfrak{A} \supset \mathfrak{A}a\mathfrak{A}$. Furthermore $\mathfrak{A} \supset \mathfrak{A}\mathfrak{A}$. Combining we have $\mathfrak{A} \supset \mathfrak{A}a\mathfrak{A}$ and $\mathfrak{A}m \supset \mathfrak{A}a\mathfrak{A}m$ for any m in \mathfrak{M} . However, since there is but one coset on the left, we must have equality. Thus the cosets containing elements of \mathfrak{A} are left scalar units. From Theorem 2 if m and m' lie in the same coset there is an a in \mathfrak{A} such that $am \supset m'$. Hence the homomorphism is a strong left unit homomorphism which has as fundamental units those cosets containing elements of \mathfrak{A} .

Conversely, let $\{e^*_i\}$ be a set of fundamental units of \mathfrak{M}^* and \mathfrak{A} the totality of elements of \mathfrak{M} corresponding to them. Suppose $\mathfrak{A}\mathfrak{A} \supset b$, where

¹⁰ For a complete discussion of units cf. D. and O., pp. 710-714.

¹¹ This is an improvement of D. and O., Theorem 13. p. 721.

b is not in \mathfrak{A} . Then $\{e^*_i\}\{e^*_i\} \supset b^*$, where b^* is not in $\{e^*_i\}$. But $\{e^*_i\}\{e^*_i\} = \{e^*_i\}$, which yields a contradiction. Thus \mathfrak{A} is a subsystem. If for some a in \mathfrak{A} $am \supset m'$, then m and m' have the same image m^* . Since the homomorphism is a strong left unit homomorphism, there is an a' in \mathfrak{A} such that $a'm' \supset m$. Hence \mathfrak{A} is left reversible. But the same argument shows that two elements lying in the same coset of \mathfrak{A} have the same image and two elements with the same image lie in the same coset. Therefore \mathfrak{M}^* is isomorphic to the m -system of the left coset expansion $\mathfrak{M}/\mathfrak{A}$.

COROLLARY. If \mathfrak{M} is a multigroup then \mathfrak{M}^* is a multigroup and \mathfrak{A} is a right submultigroup (that is, in \mathfrak{A} the second relation of Axiom 3 is satisfied).

Let us now introduce an extremely important class of subsystems, those which we shall call left normal subsystems. A subsystem \mathfrak{A} of an m -system \mathfrak{M} is left normal in \mathfrak{M} if for any m in \mathfrak{M} we have $\mathfrak{A}m \supset m\mathfrak{A}$.¹² With this type of subsystem we may associate a homomorphism in which we impose a more rigid condition on the inverse correspondence than we assumed in the previous theorem. We say that a homomorphism \mathfrak{M} to \mathfrak{M}^* is a strong left homomorphism if whenever in \mathfrak{M}^* $m^*_i m^*_j \supset m^*_k$ there is for every m_j and m_k in \mathfrak{M} corresponding to m^*_j and m^*_k respectively some m_i corresponding to m^*_i such that $m_i m_j \supset m_k$.¹³ We then have the following theorem.¹⁴

THEOREM 4. Let \mathfrak{A} be a left reversible, left normal subsystem of an m -system \mathfrak{M} . Then \mathfrak{M} is strongly left homomorphic to the m -system of the left coset expansion $\mathfrak{M}/\mathfrak{A}$ and $\mathfrak{M}/\mathfrak{A}$ has as an absolute unit \mathfrak{A} . Conversely by any strong left homomorphism $\mathfrak{M} \rightarrow \mathfrak{M}^*$ wherein \mathfrak{M}^* contains an absolute unit e^* , those elements of \mathfrak{M} which correspond to e^* form a left reversible, left normal subsystem \mathfrak{A} of \mathfrak{M} such that the m -system of the left coset expansion $\mathfrak{M}/\mathfrak{A}$ is isomorphic to \mathfrak{M}^* .

Proof. Let $\mathfrak{A}m_1\mathfrak{A}m_2 \supset \mathfrak{A}m_3$. If $r \in \mathfrak{A}m_2$, $s \in \mathfrak{A}m_3$ we have to find an x in $\mathfrak{A}m_1$ such that $xr \supset s$. Since any element in a coset generates the coset we may write $\mathfrak{A}m_1\mathfrak{A}r \supset s$. From left normality, $\mathfrak{A}\mathfrak{A}m_1r \supset s$. Therefore $\mathfrak{A}m_1r \supset s$. Hence for some a in \mathfrak{A} , $am_1r \supset s$ and for some x in $\mathfrak{A}m_1$, $xr \supset s$. We thus have a strong left homomorphism. From the preceding theorem we know that any coset containing an element of \mathfrak{A} is a left scalar unit. But for any m in \mathfrak{M} and any a in \mathfrak{A} we have $\mathfrak{A}m \supset \mathfrak{A}\mathfrak{A}m \supset \mathfrak{A}m\mathfrak{A} \supset \mathfrak{A}m\mathfrak{A}$. Since there is but one coset on the left we must have equality and $\mathfrak{A}a$ is a

¹² I have learned by correspondence that M. Krasner has also used this type of normality.

¹³ This type of homomorphism was introduced in D. and O., p. 721.

¹⁴ The same theorem for normal, reversible submultigroups is proved in D. and O., Theorem 1, p. 724.

right scalar unit. It is thus an absolute unit and since there can be but one absolute unit in any m -system, $\mathfrak{A}a = \mathfrak{A}$.

To show the converse we need only, by virtue of the preceding theorem, prove that \mathfrak{A} is left normal. Consider any $r \subset ma$. We know r and m have the same image since e^* is an absolute unit. From the strong left homomorphism there exists an a' such that $a'm \supset r$ and hence \mathfrak{A} is left normal.

COROLLARY. *If \mathfrak{M} is a multigroup, then \mathfrak{A} is a submultigroup.*

6. Conjugate subsystems. In seeking an analogue to strong normality in multigroups¹⁵ we are led to the following notion. A subsystem \mathfrak{A} of an m -system \mathfrak{M} is *left scalic* if for any m_i and m_j in \mathfrak{M} there exists an m'_j such that $\mathfrak{A}m'_j \supset m_i\mathfrak{A}m_j$. We cannot show, as in the case of multigroups, that a left scalic subsystem satisfies a normality condition. However we may show that if it is left reversible its cosets form a scalar m -system, where by *scalar m -system* we mean an m -system in which multiplication is unique; that is, the product of any two elements of the system is a single element.

THEOREM 5. *The necessary and sufficient condition that an m -system \mathfrak{M} be strongly left unit homomorphic to a scalar m -system \mathfrak{M}^* is that the m -system of the coset decomposition of \mathfrak{M} with respect to some left reversible, left scalic subsystem \mathfrak{A} is isomorphic to \mathfrak{M}^* .*

Proof. In view of Theorem 3, it is only necessary to show that if \mathfrak{A} is left scalic, $\mathfrak{M}/\mathfrak{A}$ is a scalar m -system; and if \mathfrak{M}^* is a scalar m -system, \mathfrak{A} is left scalic. Since for any m_i and m_j there is an m'_j such that $\mathfrak{A}m'_j \supset m_i\mathfrak{A}m_j$, we must have $\mathfrak{A}m'_j \supset \mathfrak{A}m_i\mathfrak{A}m_j$. As there is but a single coset on the left, we then have equality, and hence the product of any two cosets is a single coset. Conversely if for any $\mathfrak{A}m_i$ and $\mathfrak{A}m_j$ we have $\mathfrak{A}m'_j = \mathfrak{A}m_i\mathfrak{A}m_j$ we then must have $\mathfrak{A}m'_j \supset m_i\mathfrak{A}m_j$ since \mathfrak{A} is left reversible. Thus \mathfrak{A} is left scalic.

The concept of left scalic subsystems leads naturally to the notion of left conjugate subsystems. Two subsystems \mathfrak{A} , \mathfrak{B} are *left conjugates* if there exist some m and some m' such that for any m_j there exist m'_j and m''_j which satisfy the following relations:

$$\mathfrak{A}m'_j \supset m\mathfrak{B}m_j \quad \mathfrak{B}m_j \supset m'\mathfrak{A}m'_j \quad \mathfrak{B}m''_j \supset m'\mathfrak{A}m_j \quad \mathfrak{A}m_j \supset m\mathfrak{B}m''_j.$$

It is obvious that if the m -system considered is a group then the above construct is the ordinary conjugate of a group.

THEOREM 6. *Left conjugate is a symmetric, reflexive, transitive relation.*

Proof. The symmetry is obvious. The reflexivity follows from the fact that \mathfrak{A} is a subsystem, for we may choose m and m' as elements in \mathfrak{A} . Then if we choose m'_j and m''_j equal to m_j we find that we may take $\mathfrak{B} = \mathfrak{A}$ in the

¹⁵ A strongly normal submultigroup is one whose coset expansion forms a group; cf. D. and O., p. 728 ff.

definition of left conjugate. To establish the transitivity assume that we have given subsystems \mathfrak{A} , \mathfrak{B} , and \mathfrak{C} satisfying

$$(1) \quad \mathfrak{A}m_j' \supset m\mathfrak{B}m_j \quad \mathfrak{B}m_j \supset m'\mathfrak{A}m_j' \quad \mathfrak{B}n_i' \supset n\mathfrak{C}n_i \quad \mathfrak{C}n_i \supset n'\mathfrak{B}n_i'$$

$$(2) \quad \mathfrak{B}m_j'' \supset m'\mathfrak{A}m_j \quad \mathfrak{A}m_j \supset m\mathfrak{B}m_j'' \quad \mathfrak{C}n_i'' \supset n'\mathfrak{B}n_i \quad \mathfrak{B}n_i \supset n\mathfrak{C}n_i''.$$

Select $m_j = n_i'$ in (1). Then

$$\mathfrak{A}m_j' \supset m\mathfrak{B}n_i' \supset mn\mathfrak{C}n_i \supset r\mathfrak{C}n_i, \quad r \subset mn;$$

$$\mathfrak{C}n_i \supset n'\mathfrak{B}n_i' \supset n'm'\mathfrak{A}m_j' \supset r'\mathfrak{A}m_j', \quad r' \subset n'm'.$$

Select $n_i = m_j''$ in (2). Then

$$\mathfrak{C}n_i'' \supset n'\mathfrak{B}m_j'' \supset n'm'\mathfrak{A}m_j \supset r'\mathfrak{A}m_j; \quad \mathfrak{A}m_j \supset m\mathfrak{B}m_j'' \supset mn\mathfrak{C}n_i'' \supset r\mathfrak{C}n_i''.$$

Since m_j is arbitrary we may replace it by the n_i of the preceding relation. We then see that \mathfrak{A} and \mathfrak{C} are left conjugates, for note that r and r' are independent of m_j and n_i .

THEOREM 7. *A left reversible, left scalic submultigroup¹⁶ is its own only left reversible left conjugate.*

Proof. Suppose \mathfrak{A} is left reversible and left scalic and there exists a left reversible subsystem \mathfrak{B} such that

$$(1) \quad \mathfrak{A}m_j' \supset m\mathfrak{B}m_j$$

$$(2) \quad \mathfrak{B}m_j \supset m'\mathfrak{A}m_j'$$

$$(1') \quad \mathfrak{B}m_j'' \supset m'\mathfrak{A}m_j$$

$$(2') \quad \mathfrak{A}m_j \supset m\mathfrak{B}m_j''.$$

It is easy to show that \mathfrak{A} is a normal submultigroup.¹⁷ But this, together with (2), yields $\mathfrak{B}m_j \supset \mathfrak{A}r$ for any m_j and some r depending on m_j . Choose $m_j = b$. Then $\mathfrak{B} \supset \mathfrak{A}b$. If we multiply through by \mathfrak{B} on the right, we obtain $\mathfrak{B} \supset \mathfrak{A}\mathfrak{B} = \mathfrak{B}\mathfrak{A} \supset \mathfrak{A}$. We now establish the reverse inequality. From (2') and (1') we have $\mathfrak{A}m_j \supset \mathfrak{A}mm'\mathfrak{A}m_j$. As there is but one coset on the left we must have equality and mm' is in \mathfrak{A} since m_j is arbitrary and \mathfrak{A} is the only left scalar unit. In (2) choose $m_j = m_j''$ and multiply through by m on the left. Then from (2') $\mathfrak{A}m_j \supset \mathfrak{A}m_j'$. Hence there are m_j in (2) such that m_j' is arbitrary. Choose s so that $sm \supset b$ for some b in \mathfrak{B} . From (1), multiplying through by s on the left, we have $s\mathfrak{A}m_j' \supset \mathfrak{B}m_j$. Hence $\mathfrak{A}s\mathfrak{A}m_j' \supset \mathfrak{A}\mathfrak{B}m_j$. Since \mathfrak{A} is left scalic the left side is a single arbitrary coset. We may choose that coset equal to \mathfrak{A} . Then $\mathfrak{A} \supset \mathfrak{B}m_j\mathfrak{A}$ for some m_j . But in a two sided coset expansion any coset contains its generating element,¹⁸ and so m_j is in \mathfrak{A} . Hence $\mathfrak{A} \supset \mathfrak{B}\mathfrak{A} = \mathfrak{A}\mathfrak{B} \supset \mathfrak{B}$. Since we have already established the reverse inequality, $\mathfrak{A} = \mathfrak{B}$.

We define the *left normalizer* of a left reversible subsystem \mathfrak{A} of an

¹⁶ We may readily show that such a submultigroup is strongly normal in the sense defined in the paper by Eaton and Ore, "Remarks on multigroups," *American Journal of Mathematics*, this number, pp. 67-71.

¹⁷ Cf. Eaton and Ore, *op. cit.*

¹⁸ Cf. D. and O. Theorem 10, p. 718.

m -system \mathfrak{M} to be the totality of elements n_i in \mathfrak{M} such that for each n_i there is an n_i' for which for any m in \mathfrak{M} there is an m' and an m'' which satisfy the following relations:

$$\mathfrak{A}m' \supset n_i \mathfrak{A}m \quad \mathfrak{A}m \supset n_i' \mathfrak{A}m' \quad \mathfrak{A}m'' \supset n_i' \mathfrak{A}m \quad \mathfrak{A}m \supset n_i \mathfrak{A}m''.$$

THEOREM 8. *The left normalizer \mathfrak{N} of a left reversible subsystem \mathfrak{A} is a left reversible subsystem in which \mathfrak{A} is left scalar.*

Proof. Let n_1 and n_2 be two elements of \mathfrak{N} . Then from the symmetry of the definition of left normalizer, n_1' and n_2' are in \mathfrak{N} . Let u be any element in the product $n_1 n_2$ and select v as some element in the product $n_2' n_1'$. In the definition of left normalizer let m_1 (the arbitrary m associated with n_1) equal m_2' . We then obtain $\mathfrak{A}m_1' \supset u \mathfrak{A}m_2$ and $\mathfrak{A}m_2 \supset v \mathfrak{A}m_1'$. If we now select $m_2 = m_1''$ we find $\mathfrak{A}m_2'' \supset v \mathfrak{A}m_1$ and $\mathfrak{A}m_1 \supset u \mathfrak{A}m_2''$. Since in the derived relations both m_1 and m_2 are arbitrary, u is in \mathfrak{N} and \mathfrak{N} is a subsystem. Suppose $nr \supset s$ for some n in \mathfrak{N} . Take $m = r$ in the definition. Since \mathfrak{A} is left reversible we may take $m' = s$. Then we have $\mathfrak{A}r \supset n' \mathfrak{A}s$. If we multiply through by \mathfrak{A} on the left and observe that the left hand side is a single coset we have $\mathfrak{A}r = \mathfrak{A}n' \mathfrak{A}s$. Then for some a , $an's \supset r$. But all the elements of \mathfrak{A} are in \mathfrak{N} and hence every element in the product an' is in \mathfrak{N} . But for some n_1 in this product we must have $n_1 s \supset r$. Therefore \mathfrak{N} is left reversible. That \mathfrak{A} is left scalar in \mathfrak{N} follows immediately from the definition of left scalar.

7. Structure properties of normal subsystems. Let us in this section derive certain structure theorems for left normal, left reversible subsystems of an m -system which will permit us to formulate a Jordan-Hölder theorem for these subsystems.¹⁹

THEOREM 9. *Let \mathfrak{M} be an m -system and \mathfrak{A} and \mathfrak{B} left reversible, left normal subsystems. Then the crosscut $(\mathfrak{A}, \mathfrak{B})$ is a non-void subsystem which is left reversible and left normal in both \mathfrak{A} and \mathfrak{B} .*

Proof. $\mathfrak{D} = (\mathfrak{A}, \mathfrak{B})$ is non-void since, from left normality, $\mathfrak{A}\mathfrak{B} \supset \mathfrak{B}\mathfrak{A} \supset a'$. Hence there is an a and a b such that $ab \supset a'$. From left reversibility there is an a'' such that $a''a' \supset b$. Since \mathfrak{A} is a subsystem b lies in \mathfrak{A} and hence in \mathfrak{D} . \mathfrak{D} is a subsystem since $\mathfrak{D}\mathfrak{D}$ lies in both \mathfrak{A} and \mathfrak{B} and hence in \mathfrak{D} . Suppose for some d , a_1 , and a_2 we have $da_1 \supset a_2$. Then for some b' , $b'a_2 \supset a_1$. But $\mathfrak{A}b' \supset b'\mathfrak{A}$. This implies that for some a' in \mathfrak{A} , $a'b' \supset a_1$. From left reversibility $a''a_1 \supset b'$ and as before b' lies in \mathfrak{A} and hence in \mathfrak{D} . Thus \mathfrak{D} is left reversible in \mathfrak{A} and similarly in \mathfrak{B} . To show left normality, consider any $a_1 \subset a\mathfrak{D}$. Then for some b , $ab \supset a_1$. Since $\mathfrak{B}a \supset a\mathfrak{B}$ there is a b' such that

¹⁹ All the results contained in this section have been previously proved for normal, reversible submultigroups in D. and O., pp. 725-727. Our extension to left normal, left reversible subsystems is, of course, equally valid when the m -system is a multigroup.

$b'a \supset a_1$. The left normality of \mathfrak{D} in \mathfrak{A} follows immediately if we can show that b' lies in \mathfrak{A} . We know $\mathfrak{A}b' \supset b'\mathfrak{A}$ and thus there is an a' such that $a'b' \supset a_1$. Then for some a'' , $a''a_1 \supset b'$ and as before b' lies in \mathfrak{A} and hence in \mathfrak{D} . Similarly \mathfrak{D} is left normal in \mathfrak{B} .

THEOREM 10. *Let \mathfrak{M} be an m -system and \mathfrak{A} and \mathfrak{B} left reversible, left normal subsystems. Then is the union $[\mathfrak{A}, \mathfrak{B}]$ a left reversible, left normal subsystem and $[\mathfrak{A}, \mathfrak{B}] = \mathfrak{AB}$.*

Proof. $\mathfrak{AB} = \mathfrak{AABB} \supset \mathfrak{ABAB}$. Hence \mathfrak{AB} is a subsystem. But $\mathfrak{AB} \supset \mathfrak{B}$. Also $\mathfrak{AB} \supset \mathfrak{BA} \supset \mathfrak{A}$. Since $[\mathfrak{A}, \mathfrak{B}]$ is the least subsystem containing both \mathfrak{A} and \mathfrak{B} we must have $[\mathfrak{A}, \mathfrak{B}] = \mathfrak{AB}$. Furthermore \mathfrak{AB} is left reversible, for suppose $rm \supset m'$ where $ab \supset r$. Then $abm \supset m'$. Hence there is an $x \subset bm$ such that $ax \supset m'$. There is then an a' and a b' such that $a'm' \supset x$ and $b'x \supset m$. Hence there exists an $s \subset b'a'$ such that $sm' \supset m$. But $\mathfrak{AB} = \mathfrak{BA}$ since they are both equal to $[\mathfrak{A}, \mathfrak{B}]$. Thus s is in \mathfrak{AB} and \mathfrak{AB} is left reversible. Finally \mathfrak{AB} is left normal since $\mathfrak{AB}m \supset \mathfrak{AmB} \supset m\mathfrak{AB}$.

THEOREM 11. *The Dedekind Relation. Let \mathfrak{M} be an m -system and \mathfrak{A} , \mathfrak{B} , and \mathfrak{C} left reversible, left normal subsystems such that $\mathfrak{C} \supset \mathfrak{A}$. Then*

$$(\mathfrak{C}, [\mathfrak{A}, \mathfrak{B}]) = [\mathfrak{A}, (\mathfrak{C}, \mathfrak{B})].$$

Proof. Let s be in $[\mathfrak{A}, (\mathfrak{C}, \mathfrak{B})]$. Then s is in \mathfrak{AC} and hence in \mathfrak{C} and also in \mathfrak{AB} . Thus s is in $(\mathfrak{C}, [\mathfrak{A}, \mathfrak{B}])$. Any element in $(\mathfrak{C}, [\mathfrak{A}, \mathfrak{B}])$ is in \mathfrak{C} . Let c be such an element. We then have for some a and some b , $ab \supset c$. From left reversibility there is an a' such that $a'c \supset b$. Since a' is in \mathfrak{C} so also is b . Thus c is in both \mathfrak{AB} and \mathfrak{AC} and is consequently in $[\mathfrak{A}, (\mathfrak{C}, \mathfrak{B})]$.

THEOREM 12. *The Isomorphism Theorem. Let \mathfrak{M} be an m -system and \mathfrak{A} and \mathfrak{B} left reversible, left normal subsystems. Let X/Y denote the m -system of the left coset expansion of X with respect to Y . Then*

$$[\mathfrak{A}, \mathfrak{B}]/\mathfrak{A} \cong \mathfrak{B}/(\mathfrak{A}, \mathfrak{B}).$$

Proof. Let $(\mathfrak{A}, \mathfrak{B}) = \mathfrak{D}$ and $[\mathfrak{A}, \mathfrak{B}] = \mathfrak{AB} = \mathfrak{N}$. Let

$$(1) \quad \mathfrak{B} = \mathfrak{D} + \mathfrak{Db}_2 + \cdots + \mathfrak{Db}_i + \cdots$$

be the left coset decomposition of \mathfrak{B} with respect to \mathfrak{D} . Then

$$(2) \quad \mathfrak{N} = \mathfrak{A} + \mathfrak{Ab}_2 + \cdots + \mathfrak{Ab}_i + \cdots$$

is the left coset decomposition of \mathfrak{N} with respect to \mathfrak{A} for firstly every element of \mathfrak{N} lies in some coset of (2) since every element of \mathfrak{B} lies in some coset of (1) and $\mathfrak{N} = \mathfrak{AD}$. Further the cosets of (2) are distinct since if $ab_i \supset b_j$ then from left normality $b'a \supset b_j$ and from left reversibility a is in \mathfrak{B} and hence in \mathfrak{D} . But we would then have a contradiction of the assumption that (1) is a left coset decomposition. Now let $\mathfrak{Ab}_i\mathfrak{Ab}_j \supset \mathfrak{Ab}_k$. Then from left normality $\mathfrak{Ab}_ib_j \supset \mathfrak{Ab}_k$ and for some a , $ab_ib_j \supset b_k$. But as before we must then have

a lies in \mathfrak{B} and hence in \mathfrak{D} . This implies $\mathfrak{D}b_i\mathfrak{D}b_j \supset \mathfrak{D}b_k$. Conversely if $\mathfrak{D}b_i\mathfrak{D}b_j \supset \mathfrak{D}b_k$ then for some a and a' , $ab_ia'b_j \supset b_k$ and $\mathfrak{A}b_i\mathfrak{A}b_j \supset \mathfrak{A}b_k$. The isomorphism is thus established.

The preceding theorems are sufficient to prove an analogue of the Jordan-Hölder theorem for m -systems. A chain of subsystems $\mathfrak{A} \supset \mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \cdots \supset \mathfrak{A}_n = \mathfrak{B}$ is a *left composition series* between \mathfrak{A} and \mathfrak{B} when each \mathfrak{A}_i is a left reversible, left normal subsystem in the preceding and when no other terms can be intercalated in the chain. We then have:

THEOREM 13. *Any two left composition series between \mathfrak{A} and \mathfrak{B} have the same length and the m -systems of the left coset expansion of consecutive terms in one series are isomorphic in some order to those in the other series.*

The proof of the theorem is by induction on the length of the chain and is entirely similar to the proof of the corresponding theorem in group theory.

We may mention that we may define, as in group theory,²⁰ a *left quasi-normal* subsystem \mathfrak{A} of an m -system \mathfrak{M} to be such that $\mathfrak{A}\mathfrak{B} \supset \mathfrak{B}\mathfrak{A}$ for every subsystem \mathfrak{B} of \mathfrak{M} . It is then possible to formulate a Jordan-Hölder theorem involving strong structure isomorphism for the left quasi-normal, reversible subsystems.

8. Coset decomposition of groups. We may use the preceding results to characterize completely the coset decomposition of a group with respect to any subgroup.

THEOREM 14. *The necessary and sufficient condition that a partition of a group \mathfrak{G} into disjoint subsets be the left coset decomposition with respect to a subgroup \mathfrak{S} is that the m -system \mathfrak{M} of the partition be such that*

- (i) \mathfrak{M} contains a left scalar unit E ;
- (ii) $XA \supset A$ implies $X = E$.

Proof. Necessary. \mathfrak{S} is obviously a left scalar unit of \mathfrak{M} . Suppose $\mathfrak{S}g_i\mathfrak{S}g_j \supset \mathfrak{S}g_j$. Then for some h and h' in \mathfrak{S} , $hg_ih'g_j = g_j$. This implies g_i is in \mathfrak{S} and $\mathfrak{S}g_i = \mathfrak{S}$.

Sufficient. If \mathfrak{G} is finite the elements corresponding to E obviously form a subgroup. If \mathfrak{G} is not finite we still must have that e , the unit of \mathfrak{G} , corresponds to E since E is the only left unit. Furthermore if the inverse of some element corresponding to E corresponds to X , we must have $XE \supset E$ and hence $X = E$. Now let r and r' be two elements of \mathfrak{G} with the same image R . We may determine x in \mathfrak{G} such that $xr = r'$. Then $XR \supset R$ and $X = E$. Thus the homomorphism of \mathfrak{G} to \mathfrak{M} is a strong left unit homomorphism and by Theorem 3 \mathfrak{M} is isomorphic to the m -system of the left coset expansion of \mathfrak{G} with respect to the subgroup corresponding to E .

YALE UNIVERSITY.

²⁰ Cf. Oystein Ore, *op. cit.*, p. 162.

A NEW PROOF FOR A METRICALLY TRANSITIVE SYSTEM.*

By GUSTAV A. HEDLUND.

1. Introduction. Two distinct methods have been used to prove that the flows defined by the geodesics on suitably restricted surfaces of constant negative curvature are metrically transitive. The first of these [2, 3]¹ involves the use of symbolism to characterize the geodesics and the proof is restricted to those surfaces for which a suitable symbolism has been devised. The second of these methods [6, 7, 8] makes use of the theory of harmonic functions and is valid for all complete surfaces of constant negative curvature and of finite area. It is the opinion of the author that both of these methods involve excessive machinery and it would seem desirable to derive a more simple and straightforward proof of the result under discussion.

The present paper gives a new method of proof of the metric transitivity of the flow defined by the geodesics on any closed orientable surface of constant negative curvature. It seems to the writer that the present proof is considerably simpler than any previously given. The method extends readily to the general class of complete surfaces of constant negative curvature and of finite area.

2. Two-dimensional manifolds of constant negative curvature. Let Ψ be the interior of the unit circle U , $x^2 + y^2 = 1$. To Ψ we assign the metric

$$(2.1) \quad ds^2 = \frac{4(dx^2 + dy^2)}{c(1 - x^2 - y^2)^2}, \quad c > 0,$$

the Gaussian curvature of which is $-c$. The metric (2.1) assigns a length to curves in Ψ and this length is termed *hyperbolic length* or *H-length*. Angle is euclidean angle and the element of (hyperbolic) area is

$$(2.2) \quad d\sigma = \frac{4dxdy}{c(1 - x^2 - y^2)^2}.$$

The geodesics defined by (2.1) are arcs of circles orthogonal to U and are called *hyperbolic lines* or *H-lines*. An *H-line* is uniquely determined by two points of U and these points are the *points at infinity* of the *H-line*. The *hyperbolic distance* or *H-distance* between two points of Ψ is defined to be the *H-length* of the unique *H-line* segment joining the points.

* Received November 17, 1939.

¹The numbers in brackets refer to the bibliography.

A *horocycle* is a euclidean circle internally tangent to U . The point A of contact of the horocycle with U is the *point at infinity* of the horocycle. Let \bar{r} denote the minimum H -distance from the origin O to an arbitrary point of the horocycle, and let $r = +\bar{r}$ or $-\bar{r}$ according as O is interior or exterior to the horocycle. The point at infinity A and the constant r uniquely determine a horocycle and this horocycle will be denoted by $C(A, r)$. The horocycle $C(A, r)$ is an orthogonal trajectory of the set of H -lines having A as one point at infinity.

The metric (2.1) is invariant under linear fractional transformations which take Ψ into Ψ , so that under such transformations, hyperbolic distance, angle and area are invariant.

Let F be a Fuchsian group with U as principal circle (cf. [1], Ch. III). Such a group has a normal fundamental region containing the origin and bounded by H -lines or H -line segments which are congruent in pairs. If to this domain is added a suitably chosen subset of the vertices and a suitably chosen subset of the sides, the resulting region R is such that no two points of it are congruent and any point of Ψ is congruent to some point of R . We assume that F is such that the closure of R contains no points of U . It follows that F is of the first kind and has a finite set of generators, while the region R has a finite set of sides.

If points which are congruent under F are considered identical there is defined a closed two-dimensional manifold $M(F, c)$ of constant negative curvature. In the case in which $M(F, c)$ contains no singular points, all the transformations of F are hyperbolic. The genus of $M(F, c)$ is then necessarily greater than one.

An *element* e in Ψ is a point P of Ψ together with a direction at that point and can be specified by three coördinates (x, y, ϕ) , where x and y are the coördinates of P and ϕ , $0 \leq \phi < 2\pi$, is an angular coördinate measured positively in the counterclockwise sense from the directed H -ray which has P as initial point and is part of the directed H -line which passes through P and has $(1, 0)$ as initial point at infinity. The point P is the point *bearing* the element e . A neighborhood of the element (x_1, y_1, ϕ_1) is the set (x, y, ϕ) such that

$$H(P, P_1) < \delta, \quad \|\phi - \phi_1\| < \delta,$$

where P is the point (x, y) , P_1 is the point (x_1, y_1) , $H(P, P_1)$ denotes the H -distance between P and P_1 , $\|\phi - \phi_1\|$ denotes the least value of the set $|\phi - \phi_1 + 2n\pi|$, ($n = 0, \pm 1, \pm 2, \dots$), and $\delta > 0$. Let \mathcal{E} denote the space of elements in Ψ with neighborhoods thus defined.

A transformation of F carries an element into a *congruent* element. The

space Ω of elements on $M(F, c)$ is the space obtained by identifying congruent elements of \mathcal{E} . If neighborhoods in Ω are defined by correspondence with the neighborhoods defined in \mathcal{E} (cf. [9], p. 32), Ω is a Hausdorff space. It is easily shown that Ω is separable and regular and it follows that a metric yielding an equivalent topology can be assigned to Ω . With such a metric assigned, the diameter of a subset of Ω is defined.

An element (x, y, ϕ) determines a unique point of Ω , and this point will be called either the *point of Ω determined by (x, y, ϕ)* or simply the *point (x, y, ϕ) of Ω* .

If measure is defined in \mathcal{E} by means of the volume element

$$d\sigma d\phi,$$

where $d\sigma$ is given by (2.1), congruent measurable sets of \mathcal{E} have the same measure and this measure serves to define measure in Ω (cf., e. g., [4]). The hyperbolic lines or geodesics define a *geodesic flow* G_s in Ω (cf. [4]) and G_s is a measure preserving transformation of Ω into itself which is defined for each real s .

3. The tubular property of invariant sets. Let A be a point of U and let α be the smallest positive angle which the radius OA forms with the positive x -axis. The points of U are then in 1 — 1 correspondence with the interval $0 \leq \alpha < 2\pi$ and the horocycle $C(A, r)$ can be denoted by $C(\alpha, r)$.

Let $C(\alpha, r)$ be a horocycle, some point of which is in the region R . The points of $C(\alpha, r)$ in R form one or more connected arcs of $C(\alpha, r)$. Let this set of arcs be denoted by $a(\alpha, r)$ and let the number of arcs in this set be $N(\alpha, r)$. Since a horocycle can cross an H -line in at most two points, the number $N(\alpha, r)$ is not greater than twice the number of sides of R . Since R is determined by F , and since, under the assumptions we have made, R has a finite number of sides, there exists a uniform upper bound N_a for the integers $N(\alpha, r)$, where this upper bound is determined by F . Let $L(\alpha, r)$ denote the H -length of the shortest arc of $C(\alpha, r)$ which contains all the arcs $a(\alpha, r)$. Since the closure of R contains no points of U , there is a uniform upper bound L_a for the numbers $L(\alpha, r)$ and again L_a is completely determined by F .

Consider the set of arcs $a(\alpha, r)$ and the elements externally normal to $C(\alpha, r)$ at the points of $a(\alpha, r)$. The points of Ω which correspond to these elements form a set of arcs $a_e(\alpha, r)$ of Ω and the totality of arcs $a_e(\alpha, r)$ obtained by considering all admissible values of α and r form a set which is identical with Ω . By considering elements internally normal to $C(\alpha, r)$ we obtain similarly a division of Ω into arcs $a_i(\alpha, r)$.

Since the closure of R lies interior to U , the values of r assumed in the

determination of the arcs $a_e(\alpha, r)$ or $a_i(\alpha, r)$ lie between two suitably chosen constants r_1 and r_2 such that $r_1 < 0 < r_2$. It follows that the values of (α, r) determining arcs $a_e(\alpha, r)\{a_i(\alpha, r)\}$ form a set $\mathcal{A}_e\{\mathcal{A}_i\}$ of the rectangle \mathcal{R} , $0 \leq \alpha < 2\pi$, $r_1 < r < r_2$.

Let the rectangle \mathcal{R} be divided into a net $\bar{\Delta}_n$ of n^2 rectangles by n lines parallel to the α -axis and n lines parallel to the r -axis. We assume that the sequence of nets $\bar{\Delta}_1, \bar{\Delta}_2, \dots$ has been so determined that the maximum diameter (measured in terms of euclidean distance in \mathcal{R}) of the rectangles of the net $\bar{\Delta}_n$ approaches zero as n becomes infinite. It is to be understood that any one of the rectangles of the net $\bar{\Delta}_n$ includes just one vertex, namely the lower left corner, and two open sides, the lower horizontal and the left vertical. The division of \mathcal{R} into the net $\bar{\Delta}_n$ effects a division of the set $\mathcal{A}_e\{\mathcal{A}_i\}$ into subsets, a subset being the points of $\mathcal{A}_e\{\mathcal{A}_i\}$ in a rectangle of the net $\bar{\Delta}_n$. The totality of arcs $a_e(\alpha, r)\{a_i(\alpha, r)\}$ of Ω corresponding to the points of $\mathcal{A}_e\{\mathcal{A}_i\}$ in a single rectangle of $\bar{\Delta}_n$ will be called a *tube* of Ω . Thus, corresponding to the net $\bar{\Delta}_n$ there is a division of Ω into two sets of tubes and we will denote these sets by $T_e(n)$ and $T_i(n)$ respectively. A tube is evidently a measurable set and for each positive integer n , Ω is the sum of the tubes of $T_e(n)\{T_i(n)\}$.

Let E and G be measurable sets of Ω and let G be of positive measure. The relative density of the set E in the set G is defined to be the number $m(E \cdot G)/mG$.

THEOREM 3.1. *Let E be a measurable invariant set of Ω . Then, given $\epsilon > 0$, there exists a positive integer N such that if $n \gg N$, the set E , except possibly for a set of measure less than ϵ , lies in tubes of $T_e(n)\{T_i(n)\}$ in which the relative density of the set E is at least $1 - \epsilon$.*

The proof will be restricted to the case of tubes of $T_e(n)$. An analogous proof applies to the other case.

Under the measure preserving transformation G_s of Ω into itself, the set $a_e(\alpha, r)$ is transformed into a set of Ω determined by the elements externally normal to a set of segments $\sigma_s(\alpha, r)$ of the horocycle $C(\alpha, r + s)$. Let $L_s(\alpha, r)$ be the H -length of the shortest segment of $C(\alpha, r + s)$ containing the set $\sigma_s(\alpha, r)$. As $s \rightarrow -\infty$, $L_s(\alpha, r) \rightarrow 0$. Since the H -lengths $L_0(\alpha, r) = L(\alpha, r)$ are uniformly bounded by the constant L_a , the numbers $L_s(\alpha, r)$ approach zero uniformly as $s \rightarrow -\infty$. It follows that the diameter of the set of Ω determined by the elements externally normal to $C(\alpha, r + s)$ along $\sigma_s(\alpha, r)$, and hence the diameter of the set $G_s(a_e(\alpha, r))$, approaches zero as $s \rightarrow -\infty$.

Let $t_e(n)$ denote an arbitrary tube of the set $T_e(n)$. As $s \rightarrow -\infty$, the diameter of the set $G_s[t_e(n)]$ does not in general approach zero. But if n is

large, the tube $t_e(n)$ consists of elements uniformly near the elements of a set $a_e(\alpha, r)$ and if $\bar{s} < 0$ is properly chosen, the diameter of the set $G_{\bar{s}}[t_e(n)]$ will be small. Let s_n denote a value of s for which the maximum diameter of the sets $G_s[t_e(n)]$ is a minimum, s ranging over the interval $-\infty < s < 0$, while $t_e(n)$ ranges over all tubes of the set $T_e(n)$, and let this minimax diameter be $d(n)$. It is then geometrically evident that $\lim_{n \rightarrow \infty} d(n) = 0$.

Since, for any given positive integer n , the tubes $t_e(n)$ of $T_e(n)$ form a division of Ω into non-overlapping measurable subsets whose sum is Ω , the same is true of the sets $G_{s_n}[t_e(n)]$. Since G_{s_n} is a measure preserving transformation of Ω into itself and E is an invariant set of Ω , it follows that

$$m\{G_{s_n}[t_e(n)]\} = m[t_e(n)],$$

and

$$m\{G_{s_n}[t_e(n)] \cdot E\} = m[t_e(n) \cdot E].$$

Hence the relative density (if it exists) of the set E in $t_e(n)$ is identical with the relative density of E in the set $G_{s_n}[t_e(n)]$. To prove the stated theorem it suffices to prove the following lemma.

LEMMA 3.1. *Let E be a measurable set of Ω and let Δ_n , $n = 1, 2, \dots$, denote a division of Ω into a set of subsets called cells such that: (1), the number of cells in Δ_n is finite; (2) the sum of the cells in Δ_n is Ω ; (3), each of the cells forming Δ_n is measurable; and (4), if d_n denotes the maximum diameter of the cells of Δ_n , $\lim_{n \rightarrow \infty} d_n = 0$. Then given $\epsilon > 0$, there exists a positive integer N such that if $n > N$, the set E , with the exception of a set of measure less than ϵ , lies in cells of Δ_n in which the relative density of E is at least $1 - \epsilon$.*

Since E is a measurable set and $m\Omega < \infty$, corresponding to $\epsilon > 0$, there exists an open set E_0 of Ω such that $E_0 \supset E$ and $m(E_0 - E) < \epsilon^2/2$.

Let Δ_n^* denote the subset of cells of Δ_n lying in E_0 . The set Δ_n^* contains any point of E_0 which is the center of an open sphere of radius d_n made up of points of E_0 . It follows that any point of E_0 lies in the set Δ_n^* for n sufficiently large and hence, corresponding to $\epsilon > 0$, there exists a positive integer N such that $m(E_0 - \Delta_n^*) < \epsilon/2$, provided $n > N$. The following inequalities evidently hold.

$$(3.1) \quad m(\Delta_n^* - E \cdot \Delta_n^*) \leq m(E_0 - E) < \epsilon^2/2, \quad n > N.$$

$$(3.2) \quad m(E - E \cdot \Delta_n^*) \leq m(E_0 - \Delta_n^*) < \epsilon/2, \quad n > N.$$

Let $\bar{\Delta}_n^*$ denote the set of cells of Δ_n^* in which the relative density of the set E is less than $1 - \epsilon$. It follows that

$$\epsilon m \bar{\Delta}_n^* \leq m(\bar{\Delta}_n^* - \bar{\Delta}_n^* \cdot E), \quad n > N,$$

and since

$$\bar{\Delta}_n^* - \bar{\Delta}_n^* \cdot E \subset E_0 - E, \quad n > N,$$

we infer with the aid of (3.1) that

$$(3.3) \quad m \bar{\Delta}_n^* < \epsilon/2, \quad n > N.$$

Except possibly for the sum of the sets $E - E \cdot \bar{\Delta}_n^*$ and $E \cdot \bar{\Delta}_n^*$, the set E lies in cells of $\bar{\Delta}_n^*$ in which the relative density of E is at least $1 - \epsilon$. But the measure of each of the sets $E - E \cdot \bar{\Delta}_n^*$ and $E \cdot \bar{\Delta}_n^*$ is, according to (3.2) and (3.3), less than $\epsilon/2$ if $n > N$, and thus we can infer the truth of the stated lemma.

The proof of Theorem 3.1 is complete.

The following evident extension of Theorem 3.1 will be useful.

THEOREM 3.2. *Let $T_e^*(n)\{T_i^*(n)\}$ denote a division of each of the tubes of $T_e(n)\{T_i(n)\}$ into measurable non-overlapping subsets such that for each n the number of the sets in $T_e^*(n)\{T_i^*(n)\}$ is finite and their sum is Ω . Let E be a measurable invariant set of Ω . Then given $\epsilon > 0$ there exists a positive integer N such that if $n > N$, the set E , except for a set of measure less than ϵ , lies in sets of $T_e^*(n)\{T_i^*(n)\}$ in which the relative density of the set E is at least $1 - \epsilon$.*

4. Metric transitivity. Let e be the element (x, y, ϕ) . This element can also be specified by three coördinates (α, r, h) , where α and r are the numbers determining the horocycle $C(\alpha, r)$ which passes through (x, y) and has e as an exterior normal element, while h is the oriented hyperbolic arc-length on $C(\alpha, r)$, measured positively in the clockwise sense on $C(\alpha, r)$ from the point of $C(\alpha, r)$ which is nearest the origin. The transformation from (x, y, ϕ) to (α, r, h) is analytic with non-vanishing Jacobian in the set $x^2 + y^2 < 1, 0 < \phi < 2\pi$.

Let Ω^* denote the subset of Ω determined by the elements (x, y, ϕ) such that (x, y) is in the interior of the region R and $0 < \phi < 2\pi$. It is evident that $m(\Omega - \Omega^*) = 0$. If E is a measurable subset of Ω , the measure of the set $E \cdot \Omega^*$ coincides with that of E and by the transformation from (x, y, ϕ) to (α, r, h) defined above, the set $E \cdot \Omega^*$ can be represented in the (α, r, h) space by a measurable bounded set, the measure of which is defined to be that of $E \cdot \Omega$. If the metric density of E at any point of Ω^* is defined by means of cubes in the (α, r, h) space, it is well known that the metric density of E is 1 at almost all points of E and 0 at almost all points of $\Omega - E$.

LEMMA 4.1. If $e_1(\alpha_1, r_1, h_1)$ and $e_2(\alpha_1, r_1, h_2)$, $h_1 < h_2$, are elements such that the points bearing all the elements

$$(\alpha_1, r_1, h), \quad h_1 \leq h \leq h_2,$$

are interior to R and if the metric density of the measurable invariant set E is 1 at e_1 , then the metric density of E at e_2 is also 1.

We assume in the proof of this theorem that $\alpha_1 \neq 0$. If α_1 were zero, a slight rotation of the region R would permit the application of the given proof.

Under this condition the constant $\delta > 0$ can be chosen so small that all the points (α, r, h) satisfying the condition

$$|\alpha - \alpha_1| < \delta, \quad |r - r_1| < \delta, \quad h_1 - \delta \leq h \leq h_2 + \delta,$$

are in Ω^* . We denote this set by B_δ . The subsets of B_δ determined by the inequalities

$$|\alpha - \alpha_i| < \delta, \quad |r - r_1| < \delta, \quad |h - h_i| < \delta, \quad (i = 1, 2),$$

will be denoted by C_δ and D_δ , respectively. If we let

$$\frac{m(E \cdot C_\delta)}{mC_\delta} = \lambda_\delta,$$

it follows from the hypotheses of the lemma that $\lim_{\delta \rightarrow 0} \lambda_\delta = 1$. For the moment we hold δ fast.

Let $T_e(n)$ be a division of Ω into tubes as defined in § 3, and let the sets $T_e^*(n)$ be determined as follows. If a tube of $T_e(n)$ contains no points of B_δ , the tube is a set of $T_e^*(n)$, while if a tube $t_e(n)$ of $T_e(n)$ contains a point of B_δ we divide $t_e(n)$ into two sets $t_e(n) \cdot B_\delta$ and $t_e(n) - t_e(n) \cdot B_\delta$, both of which are sets of $T_e^*(n)$. The sets $T_e^*(n)$ then fulfill the conditions imposed in Theorem 3.2 and given $\epsilon > 0$, there exists a positive integer N such that if $n > N$, the set E , except possibly for a set of measure less than ϵ , lies in sets of $T_e^*(n)$ in which the relative density of E is at least $1 - \epsilon$. Let the set obtained by excluding the exceptional set from E be denoted by E_n^* . Then $m(E - E_n^*) < \epsilon$ if $n > N$, and we assume that n is so chosen.

The subsets of $T_e^*(n)$ containing points of B_δ form a set of rectangular parallelepipeds (in (α, r, h) space)

$$(4.1) \quad t_i^n, \quad (i = 1, 2, \dots, v(n)).$$

The sets

$$(4.2) \quad t_i^n \cdot C_\delta, \quad (i = 1, 2, \dots, v(n)),$$

form a division of C_δ into non-overlapping subsets (rectangular parallelepipeds) and let

$$(4.3) \quad t_{i_k}^n \cdot C_\delta, \quad (k = 1, 2, \dots, \mu(n)),$$

denote those sets of (4.2) which contain points of E_n^* . It follows that

$$(4.4) \quad m\left(\sum_{k=1}^{\mu(n)} t_{i_k}^n \cdot C_\delta\right) \geq m(E_n^* \cdot C_\delta).$$

Since

$$m(E \cdot C_\delta) = \lambda_\delta m C_\delta,$$

it follows from the condition $m(E - E_n^*) < \epsilon$ and (4.4) that

$$(4.5) \quad m\left(\sum_{k=1}^{\mu(n)} t_{i_k}^n \cdot C_\delta\right) \geq \lambda_\delta m C_\delta - \epsilon.$$

But the transformation

$$\alpha = \alpha, \quad r = r, \quad h = h + \text{const.}$$

is measure preserving in Ω (cf. [5]) and thus

$$(4.6) \quad m C_\delta = m D_\delta; \quad m(t_i^n \cdot C_\delta) = m(t_i^n \cdot D_\delta), \quad (i = 1, 2, \dots, \nu(n)).$$

We infer from (4.5) and (4.6) that

$$(4.7) \quad m\left(\sum_{k=1}^{\mu(n)} t_{i_k}^n \cdot D_\delta\right) \geq \lambda_\delta m D_\delta - \epsilon.$$

Since

$$m(t_{i_k}^n \cdot E_n^*) \geq (1 - \epsilon) m t_{i_k}^n, \quad (k = 1, 2, \dots, \mu(n)),$$

it follows that

$$(4.8) \quad m\{E_n^* \cdot (t_{i_k}^n \cdot D_\delta)\} \geq m\{t_{i_k}^n \cdot D_\delta\} - \epsilon m t_{i_k}^n, \quad (k = 1, 2, \dots, \mu(n)).$$

Summing over $k = 1, 2, \dots, \mu(n)$, we obtain

$$m(E_n^* \cdot D_\delta) \geq m\left(\sum_{k=1}^{\mu(n)} (t_{i_k}^n \cdot D_\delta)\right) - \epsilon m\left(\sum_{k=1}^{\mu(n)} t_{i_k}^n\right).$$

From this inequality and (4.7) we infer that

$$m(E_n^* \cdot D_\delta) \geq \lambda_\delta m D_\delta - \epsilon - \epsilon m B_\delta,$$

whence

$$(4.9) \quad \frac{m(E \cdot D_\delta)}{m D_\delta} \geq \frac{m(E_n^* \cdot D_\delta)}{m D_\delta} \geq \lambda_\delta - \frac{\epsilon}{m D_\delta} - \epsilon \frac{m B_\delta}{m D_\delta}.$$

Since, for a given $\delta > 0$, ϵ can be chosen arbitrarily small, we infer that

$$\frac{m(E \cdot D_\delta)}{m D_\delta} \geq \lambda_\delta.$$

This implies that the lower metric density of E at (α_1, r_1, h_2) is at least as great as the metric density of E at the point (α_1, r_1, h_1) . But the latter was assumed to be 1, and the statement of the lemma is proved.

The element obtained by rotating a given element through 180° is termed the element *opposite* the given element. The proof of the following lemma is closely analogous to that of Lemma 4.1, and will be omitted.

LEMMA 4.2. *If $e_1(\alpha_1, r_1, h_1)$ and $e_2(\alpha_1, r_1, h_2)$, $h_1 < h_2$, are elements such that the points bearing all the elements*

$$(\alpha_1, r_1, h), \quad h_1 \leq h \leq h_2,$$

are interior to R , and if the metric density of the measurable invariant set E is 1 at the element opposite e_1 , then the metric density of E at the element opposite e_2 is also 1.

THEOREM 4.1. (*Metrical Transitivity.*) *If E is a measurable invariant set of Ω , either $mE = 0$ or $m(\Omega - E) = 0$.*

It is sufficient to show that if $mE > 0$, then $mE = m\Omega$. If $mE > 0$, E contains a point $p(x, y, \phi)$, (x, y) interior to R , such that the metric density of E at p is 1. Since E is invariant under the measure preserving geodesic flow G_s , each element on the directed geodesic of which p is an element will be a point of E at which the metric density of E is 1. Since (x, y) is interior to R , there is a connected arc a_e of the horocycle C_1 of which p is a normal exterior element such that a_e contains (x, y) and lies in the interior of R . According to Lemma 4.1, each element which is externally normal to C_1 at a point of a_e is a point at which the metric density of E is unity.

Similarly, there is an arc a_i of the horocycle C_2 of which p is a normal interior element such that a_i contains (x, y) and lies in the interior of R . According to Lemma 4.2, each element which is internally normal to C_2 at a point of a_i is a point at which the metric density of E is 1.

Thus there are three possible transformations of an element such that if the metric density of E at the element is initially 1, it is 1 after the transformation. It is a simple geometrical problem to show that given any ϕ' such that $0 \leq \phi' < 2\pi$, it is possible to transform (x, y, ϕ) into (x, y, ϕ') by means of these transformations without passing out of a small neighborhood of the point (x, y) . Thus, if (x, y, ϕ) is an element at which the metric density of E is 1, then the metric density of E is 1 at all the points (x, y, ϕ') , $0 \leq \phi' < 2\pi$. If ϕ' is chosen properly, the directed geodesic determined by (x, y, ϕ') will pass through the origin, and thus some point $(0, 0, \bar{\phi})$ of Ω is a point at which the metric density of E is 1. But then all the points $(0, 0, \phi)$,

$0 \leq \phi < 2\pi$, are points at which the metric density of E is 1. By reversing this process we infer that every (x, y, ϕ) , (x, y) interior to R , $0 \leq \phi < 2\pi$, is a point at which the metric density of E is 1. It follows that $mE = m\Omega$ and the proof of the theorem is complete.

UNIVERSITY OF VIRGINIA,
CHARLOTTESVILLE, VA.

BIBLIOGRAPHY.

1. L. R. Ford, *Automorphic Functions*, New York, 1929.
2. G. A. Hedlund, "On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature," *Annals of Mathematics*, vol. 35 (1934), pp. 787-808.
3. ———, "A metrically transitive group defined by the modular group," *American Journal of Mathematics*, vol. 57 (1935), pp. 668-678.
4. ———, "The dynamics of geodesic flows," *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 241-260.
5. ———, "Fuchsian groups and mixtures," *Annals of Mathematics*, vol. 40 (1939), pp. 370-383.
6. E. Hopf, "Fuchsian groups and ergodic theory," *Transactions of the American Mathematical Society*, vol. 39 (1936), pp. 299-314.
7. ———, "Ergodentheorie," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 5 (1937).
8. ———, "Beweis des Mischungscharakters der geodätischen Strömung auf Flächen der Krümmung minus Eins und endlicher Oberfläche," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 1938, pp. 333-334.
9. H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Berlin, 1934.

THE EULER NUMBER OF A RIEMANN MANIFOLD.*

By CARL B. ALLENDOERFER.

1. Introduction. One of the chief links between the differential geometry and the topology of two dimensions is the corollary to the Gauss-Bonnet theorem which states: The integral of the total curvature of a two dimensional closed surface over the surface is equal to $2\pi N$, where N is the Euler number of the surface. Since the Gauss-Bonnet theorem is of intrinsic character, this theorem does not require the surface to be a subspace of any Euclidean space.

An alternative, but less inclusive, proof of this theorem can be given which avoids the Gauss-Bonnet theorem and uses instead the property that the surface lies in a three dimensional Euclidean space. This proof can be generalized to a closed Riemann space R_n of even dimension which is a subspace of an $n + 1$ dimensional Euclidean space, i. e., a hypersurface. In this case the theorem takes the form:

$$(1.1) \quad \int_{R_n} K dO = \frac{\omega_n}{2} N$$

where K is the total curvature of R_n , ω_n is the area of an n -sphere (a sphere whose surface is n dimensional), and N is the Euler number of R_n . We recall here that K is defined for a hypersurface as the product of the n principal curvatures, and that it can be expressed as a polynomial in the $R_{\alpha\beta\gamma\delta}$ of R_n divided by the determinant of the $g_{\alpha\beta}$. We shall not consider the case n odd, for it has been shown that (1.1) does not hold under these circumstances.¹

The problem at hand is to extend (1.1) to spaces which are not hypersurfaces. If no imbedding is to be assumed this requires a generalization of the Gauss-Bonnet theorem to more than two dimensions, and so far this has not been accomplished. Progress, however, can be made by assuming that R_n lies in a Euclidean space of $n + q$ dimensions, and on this basis we shall prove the following

THEOREM. *If a closed Riemann manifold of even dimension can be made a subspace of a Euclidean space E_{n+q} , then*

$$\int_{R_n} K dO = \frac{1}{2} \omega_n N,$$

* Received October 12, 1939.

¹ For the case of a hypersurface see H. Hopf, "Über die Curvatura integra geschlossener Hyperflächen," *Mathematische Annalen*, vol. 95 (1925), pp. 340-367.

where

$$K = \frac{R_{a_1 a_2 \beta_1 \beta_2} \cdots R_{a_{n-1} a_n \beta_{n-1} \beta_n} \epsilon^{a_1 \cdots a_n} \epsilon^{\beta_1 \cdots \beta_n}}{n! 2^{n/2} |g_{a\beta}|}.$$

The term "closed Riemann manifold" is used in the sense defined by Hopf in a paper in which the background of this problem is discussed and the present investigation suggested.²

The chief difficulty in the preparation of this paper was the definition of K since no theory of principal curvatures, etc. exists for spaces other than hypersurfaces. Instead K is defined indirectly by the use of the theory of tubes recently developed by H. Weyl.³ Once this is accomplished our theorem is an immediate application of Kronecker's index theorem⁴ and of Weyl's results.

2. Kronecker's index.⁴ An important tool in the proof is an integral theorem due to Kronecker, the proof of which is here summarized from the present point of view. Let S be an n dimensional closed Riemann manifold on which is defined a set of $n+1$ functions of class C^1 , $V^i(x)$, which satisfy $V^i V^i = 1$. By means of this set of functions we can consider a continuous mapping of S upon the unit n -sphere, Σ , whose equation is $V^i V^i = 1$. The orientations on S and Σ are those imposed by a fixed orientation in the arithmetic space of the parameters x^a . This mapping is of a definite degree d , where d is an integer, positive, negative, or zero. We seek an analytic expression for d . Consider the determinant:

$$(2.1) \quad D = \begin{vmatrix} V^1 & \cdots & V^{n+1} \\ \frac{\partial V^1}{\partial x^1} & \cdots & \frac{\partial V^{n+1}}{\partial x^1} \\ \vdots & & \vdots \\ \frac{\partial V^1}{\partial x^n} & \cdots & \frac{\partial V^{n+1}}{\partial x^n} \end{vmatrix}.$$

It is easy to show that

$$(2.2) \quad D^2 = \left| \frac{\partial V^i}{\partial x^a} \quad \frac{\partial V^i}{\partial x^b} \right|$$

² H. Hopf, "Differentialgeometrie und Topologische Gestalt," *Jahresbericht der Deutscher Math. Vereinigung*, vol. 41 (1932), pp. 209-229. Also see Hopf und Rinow, "Über den Begriff der vollständigen differentialgeometrischen Fläche," *Comm. Math. Helvet.*, vol. 3 (1931), pp. 209-225.

³ H. Weyl, "On the volume of tubes," *American Journal of Mathematics*, vol. 61 (1939), pp. 461-472. Readers should note that in Weyl's paper ω_n refers to the surface area of a sphere which incloses a volume of n dimensions. We put ω_n equal to the area of a sphere whose inclosed volume is $n+1$ dimensional.

⁴ For a full treatment see J. Tannery, *Introduction à la Théorie des Fonctions*, Note by J. Hadamard, vol. 2, pp. 437-477.

which is recognized as the determinant of the metric tensor of the n -sphere if V^i are taken as Euclidean coördinates. The area, ω_n , of the sphere is thus given by:

$$\int_S \pm \sqrt{D^2} dx^1 \cdots dx^n = d \cdot \omega_n$$

integrated over S provided the sign of the radical is chosen from point to point to allow for overlapping of the covering. This is accomplished at once by using $\int_S D dx$. For D is numerically equal to $\sqrt{D^2}$ and has a positive sign for elements on the sphere of positive orientation and a negative sign in the opposite case. Therefore $\int_S D dx = d \cdot \omega_n$.

In the special case where S is a hypersurface and where V^i is the normal vector ξ^i , we arrive at the total curvature of S . For

$$(2.3) \quad \frac{\partial \xi^i}{\partial x^\alpha} = \bar{b}_{\alpha\beta} g^{\beta\gamma} \frac{\partial y^i}{\partial x^\gamma}$$

where $\bar{b}_{\alpha\beta}$ are the negatives of the coefficients of the second fundamental form of S . From the fact that $\frac{\partial y^i}{\partial x^\alpha} \frac{\partial y^i}{\partial x^\beta} = g_{\alpha\beta}$ we have that:

$$(2.4) \quad D^2 = | \bar{b}_{\alpha\beta} g^{\beta\gamma} \bar{b}_{\gamma\delta} |;$$

or that

$$(2.5) \quad D = e \frac{| \bar{b}_{\alpha\beta} |}{| g_{\alpha\beta} |^{\frac{1}{2}}}.$$

By considering the special case

$$\xi^i = (1, 0, \cdots, 0); \quad \frac{\partial y^i}{\partial x^\alpha} = (0, \cdots, 0, 1, 0, \cdots, 0)$$

with 1 in the $(\alpha + 1)$ -th place, $\bar{b}_{\alpha\beta} = \delta_{\alpha\beta}$; $g_{\alpha\beta} = \delta_{\alpha\beta}$; it is shown that e is definitely $+1$, since (2.5) is an identity. But since K , the total curvature, equals $| \bar{b}_{\alpha\beta} | / | g_{\alpha\beta} |$, this shows that:

$$(2.6) \quad \int_S K dO = d \cdot \omega_n,$$

where $dO = \sqrt{| g_{\alpha\beta} |} dx^1 \cdots dx^n$.

Since n is even, it is necessary that N , the Euler number of S , be equal to $2d$. Hence

$$(2.7) \quad \int_S K dO = \frac{1}{2} \omega_n N$$

which is the required theorem for this special case.

3. Fundamental equations on tubes. Let $y^i = y^i(u)$ be the para-

metric equations of R_n in E_{n+q} for a neighborhood of R_n . Since it may be necessary to consider a number of sets of such parameters in order to cover R_n completely, we shall let u^a be a typical set. There now exist q mutually orthogonal unit vectors ξ_{σ^i} ($\sigma = 1 \cdots q$) which are normal to R_n . Let these be chosen as functions of class C^1 and such that the determinant $\left| \xi_{\sigma^i}, \frac{\partial y^i}{\partial x^a} \right| > 0$. The parametric equations of a tube of unit radius may then be written:

$$(3.1) \quad x^i = y^i(u) + t^{\sigma}(v)\xi_{\sigma^i}$$

where

$$(3.2) \quad t^{\sigma}t^{\sigma} = 1$$

and v^A ($A = 1 \cdots q-1$) are parameters on a $(q-1)$ sphere. Again several sets of v 's will be needed to cover the sphere. The tangent vectors of this tube are then:⁵

$$(3.3) \quad \begin{aligned} \frac{\partial x^i}{\partial u^a} &= \frac{\partial y^i}{\partial u^a} + t^{\sigma} \{ -\Omega^{\sigma}_{\alpha\beta} g^{\beta\gamma} \frac{\partial y^i}{\partial u^{\gamma}} + v_{\rho\sigma/a} \xi_{\rho^i} \} \\ \frac{\partial x^i}{\partial v^A} &= \frac{\partial t^{\sigma}}{\partial v^A} \xi_{\sigma^i}. \end{aligned}$$

The tube is moreover a hypersurface of E_{n+q} and its normal vector is then $t^{\sigma}\xi_{\sigma^i}$. This follows from (3.3) and the equations:

$$\xi_{\sigma^i} \frac{\partial y^i}{\partial u^a} = 0; \quad \xi_{\sigma^i} \xi_{\rho^i} = \delta_{\sigma\rho}; \quad t^{\sigma} \frac{\partial t^{\sigma}}{\partial v^A} = 0; \quad v_{\rho\sigma/a} = -v_{\sigma\rho/a}.$$

We then observe that:

$$(3.4) \quad \begin{aligned} \frac{\partial}{\partial u^a} (t^{\sigma}\xi_{\sigma^i}) &= -t^{\sigma}\Omega^{\sigma}_{\alpha\gamma} y_{,\delta^i} g^{\gamma\delta} + t^{\sigma} v_{\rho\sigma/a} \xi_{\rho^i}; \\ \frac{\partial}{\partial v^A} (t^{\sigma}\xi_{\sigma^i}) &= \frac{\partial t^{\sigma}}{\partial v^A} \xi_{\sigma^i}. \end{aligned}$$

Hence for the tube the D of Kronecker's index is

$$(3.5) \quad D = \begin{vmatrix} t^{\sigma}\xi_{\sigma^i} \\ \frac{\partial t^{\sigma}}{\partial v^A} \xi_{\sigma^i} \\ t^{\sigma}\bar{\Omega}^{\sigma}_{\alpha\gamma} y_{,\delta^i} g^{\gamma\delta} + t^{\sigma} v_{\rho\sigma/a} \xi_{\rho^i} \end{vmatrix} \begin{matrix} 1 \text{ row} \\ q-1 \text{ rows} \\ n \text{ rows} \end{matrix}$$

where $\bar{\Omega}^{\sigma}_{\alpha\gamma} = -\Omega^{\sigma}_{\alpha\gamma}$. And at once it follows that

$$(3.6) \quad D^2 = \left| t^{\sigma}\bar{\Omega}^{\sigma}_{\alpha\beta} t^{\rho}\bar{\Omega}^{\rho}_{\gamma\delta} g^{\beta\delta} \right| \times \left| \frac{\partial t^{\sigma}}{\partial v^A} \frac{\partial t^{\sigma}}{\partial v^B} \right|,$$

whose square root gives:

$$(3.7) \quad D = e \frac{|t^{\sigma}\bar{\Omega}^{\sigma}_{\alpha\beta}|}{|g_{\alpha\beta}|} \sqrt{t} \sqrt{g}$$

⁵ For instance see L. P. Eisenhart, *Riemannian Geometry*, p. 189, equations (56.3).

where $t = \left| \frac{\partial t^\sigma}{\partial v^A} \frac{\partial t^\sigma}{\partial v^B} \right|$. We note that \sqrt{t} is the surface element of a $q-1$ sphere, Σ . The value of e here depends essentially on the orientation chosen on R_n and on the sphere. In order to decide this matter consider the special case where: $\xi_{\sigma^i} = (0, \dots, 0, 1, 0, \dots, 0)$ with 1 in the σ -th place only; $y_{\alpha^i} = (0, \dots, 0, 1, 0, \dots, 0)$ with 1 in the $\alpha + q$ -th place only; $\bar{\Omega}^1_{\alpha\beta} = \delta_{\alpha\beta}$; $\bar{\Omega}^\sigma_{\alpha\beta} = 0$ for $\sigma \neq 1$; $g^{\gamma\delta} = \delta^{\gamma\delta}$; $v_{\rho\sigma/\alpha} = 0$. Then (3.7) becomes term by term:

$$\begin{vmatrix} t^1 & \dots & t^q \\ \frac{\partial t^1}{\partial v^{q-1}} & \dots & \frac{\partial t^q}{\partial v^{q-1}} \\ \vdots & & \vdots \\ \frac{\partial t^1}{\partial v^{q-1}} & \dots & \frac{\partial t^q}{\partial v^{q-1}} \\ \hline & & \\ & t^1 & \\ & & t^1 \\ & & \\ & & t^1 \end{vmatrix} = e \frac{(t^1)^n}{1} \sqrt{t} \sqrt{1}.$$

The value of the upper left-hand minor is $\pm \sqrt{t}$. We choose the positive orientation on Σ so that the sign of the radical is positive at all times. This shows that $e = +1$, since (3.7) is an identity. Remembering to perform all integrations in the thus determined positive sense, we have that

$$\int_{R_n, \Sigma} \frac{|t^\sigma \bar{\Omega}^\sigma_{\alpha\beta}|}{|g_{\alpha\beta}|} \sqrt{t} \sqrt{g} dv^1 \dots dv^{q-1} du^1 \dots du^n = \frac{\bar{N}}{2} \omega_{n+q-1}$$

where \bar{N} is the Euler number of the tube, provided that $n + q - 1$ is even. We have assumed that n is even, and hence require q to be odd. The case q even will be handled presently.

4. Final results. Here we follow closely the results of Weyl's recent paper in evaluating the integral on the left. Since

$$(4.1) \quad I = \int_{\Sigma} \frac{|t^\sigma \bar{\Omega}^\sigma_{\alpha\beta}|}{|g_{\alpha\beta}|} \sqrt{t} dv^1 \dots dv^{q-1}$$

is an orthogonal invariant with respect to the index σ , and since n is even, I is expressible as a polynomial in $\bar{\Omega}^\sigma_{\alpha\beta} \bar{\Omega}^\sigma_{\gamma\delta} \equiv E_{\alpha\beta/\gamma\delta}$. Following Weyl we have that:

$$(4.2) \quad I = \omega_{q-1} \frac{(n-1)(n-3) \dots 3 \cdot 1}{q(q+2) \dots (n+q-2)} K$$

where

$$K = \frac{1}{n!} \frac{E_{a_1\beta_1/a_2\beta_2} \cdots E_{a_{n-1}\beta_{n-1}/a_n\beta_n} \epsilon^{a_1 \cdots a_n} \epsilon^{\beta_1 \cdots \beta_n}}{|g_{a\beta}|}.$$

Because of the relation: $R_{a\beta\gamma\delta} = E_{a\gamma/\beta\delta} - E_{a\delta/\beta\gamma}$, we may write:

$$(4.3) \quad K = \frac{1}{2^{n/2} n!} \frac{R_{a_1 a_2 \beta_1 \beta_2} \cdots R_{a_{n-1} a_n \beta_{n-1} \beta_n} \epsilon^{a_1 \cdots a_n} \epsilon^{\beta_1 \cdots \beta_n}}{|g_{a\beta}|}.$$

Thus:

$$(4.4) \quad \int_{R_n} K \sqrt{|g_{a\beta}|} du^1 \cdots du^n \cdot \omega_{q-1} \frac{(n-1) \cdots 3 \cdot 1}{q(q+2) \cdots (n+q-2)} \\ = \frac{1}{2} \bar{N} \omega_{n+q-1}.$$

Recalling that q is odd we have that

$$\omega_{n+q-1} = \frac{(2\pi)^{n/2}}{q(q+2) \cdots (n+q-2)} \omega_{q-1} \\ \omega_n = 2 \frac{(2\pi)^{n/2}}{(n-1)(n-3) \cdots 3 \cdot 1}$$

and hence that

$$(4.5) \quad \omega_{n+q-1} = \frac{1}{2} \omega_n \left\{ \frac{(n-1)(n-3) \cdots 3 \cdot 1}{q(q+2) \cdots (n+q-2)} \omega_{q-1} \right\}.$$

Combination of (4.4) and (4.5) gives:

$$(4.6) \quad \int_{R_n} K \sqrt{|g_{a\beta}|} du^1 \cdots du^n = \frac{\bar{N}}{2} \cdot \frac{\omega_n}{2}.$$

Now we know from topology that $\bar{N} = 2N$ where N is the Euler number of R_n . For the tube is topologically the product of R_n with a $q-1$ sphere where $q-1$ is even. This leads at once to the above relation between their Euler numbers. Thus we have that for n even and q odd:

$$(4.7) \quad \int_{R_n} K \sqrt{|g_{a\beta}|} du^1 \cdots du^n \equiv \int_{R_n} K dO = \frac{N}{2} \omega_n.$$

This result extends immediately to the case q even. For if q is originally even, imbed the $n+q$ dimensional Euclidean space in a similar space of $n+q+1$ dimensions so that the parametric equations of R_n are $y^i = y^i(u)$, $i = 1 \cdots n+q$; $y^{n+q+1} = \text{constant}$. Now the proof proceeds as above, yielding the desired result, since q does not appear in the final formula whatsoever.

THE INDEX THEOREM FOR A CALCULUS OF VARIATIONS PROBLEM IN WHICH THE INTEGRAND IS DISCONTINUOUS.*

By NANCY COLE.

Introduction. The purpose of this paper is to establish Morse's Index Theorem¹ for a problem in euclidean m -space in which the integrand is discontinuous and in which the basic curve g is a broken extremal with a finite number of corners. We assume that at each corner g is cut across by a regular $(m-1)$ -manifold of class C^2 , which is not tangent to either arc of g at the corner, and that at each corner g satisfies a set of "primary incidence relations." Our integral J along g will be of the form

$$J_g = \int_{g_1} F^1(x, \dot{x}) dt + \cdots + \int_{g_n} F^k(x, \dot{x}) dt$$

where g_1, \dots, g_k indicate the extremal arcs of which g is composed. Mason and Bliss (see Bliss 1) discussed the minimizing properties of a broken extremal in a problem with a discontinuous integrand in 2-space. Miles (1) extended their results to 3-space. In order to keep the notation as simple as possible, we treat below the case $k=2$ in m -space.

1. General hypotheses. Let R be an open region in the space of the variables $(x) = (x^1, \dots, x^m)$. Let g be a simple continuous curve lying in R and composed of two successive regular arcs g_1 and g_2 , each of class C^2 . We shall represent g in the form

$$(1.1) \quad x^i = \gamma^i(t) \quad (i = 1, \dots, m)$$

where t is the arc length and increases from t' to t'' inclusive.

Let c , where $t' < c < t''$, represent the value of the parameter t at the point of intersection of g_1 and g_2 . We suppose that at the corner $t=c$, g is cut across by a regular $(m-1)$ -manifold M which is not tangent to either arc of g at $t=c$. We assume that M is representable in the form

$$(1.2) \quad x^i = z^i(\alpha^1, \dots, \alpha^n) = z^i(\alpha) \quad (i = 1, \dots, m; n = m-1)$$

where the functions $z^i(\alpha)$ are of class C^2 for (α) near (0) , and for $(\alpha) = (0)$ give the point $t=c$ on g . We term M the *deflecting manifold*.

* Received June 22, 1939.

¹ See Morse 2. Numerals following the name of an author refer to the bibliography at the end.

By a function of class C^3 in a domain S which is not entirely open, we shall mean a function of class C^3 in an open region which contains the domain S in its interior.

In the space of the variables (x) let R_1 (R_2) be a domain of points (x) near g_1 (g_2) excepting those points (x) not on M which lie on the same side of M as g_2 (g_1). Let

$$F^1(x, r) = F^1(x^1, \dots, x^m, r^1, \dots, r^m)$$

be a function of class C^3 for (x) in R_1 and (r) any set not (0) , and let

$$F^2(x, r) = F^2(x^1, \dots, x^m, r^1, \dots, r^m)$$

be a function of class C^3 for (x) in R_2 and (r) any set not (0) . We assume that each function $F^\kappa(x, r)$, $\kappa = 1$ or 2 , is positive homogeneous of order 1 in the variables (r) ; that is

$$(1.3) \quad F^\kappa(x, kr) = kF^\kappa(x, r)$$

for all numbers $k > 0$ and $(r) \neq (0)$. We assume also that the problem is positive regular along g ; that is

$$(1.4) \quad \begin{aligned} F^1_{r^i, r^j}(x, r)\lambda^i\lambda^j &> 0, \\ F^2_{r^i, r^j}(x, r)\lambda^i\lambda^j &> 0 \end{aligned} \quad (i, j = 1, \dots, m)$$

for $(x, r) = (\gamma, \dot{\gamma})$ respectively on g_1 and g_2 and for (λ) any set not (0) and not proportional to $(\dot{\gamma})$ on g_1 and g_2 respectively.

A curve of class D^1 neighboring g will be termed *admissible* if it joins the initial point t' of g to a point (z) on M and that point (z) on M to the final end point t'' of g , and crosses M just once.

For our problem the familiar integral J defined along an admissible curve γ of class D^1 neighboring g will be of the form

$$J_\gamma = \int_{\gamma_1} F^1(x, \dot{x}) dt + \int_{\gamma_2} F^2(x, \dot{x}) dt$$

where \dot{x}^i stands for the derivative of x^i with respect to the parameter t and where γ_1 and γ_2 denote the arcs of γ lying in R_1 and R_2 respectively.

We assume that the Euler equations hold along g ; that is the Euler equations

$$(1.5) \quad \begin{aligned} \frac{d}{dt} F^1_{r^i} - F^1_{x^i} &= 0, \\ \frac{d}{dt} F^2_{r^i} - F^2_{x^i} &= 0 \end{aligned} \quad (i = 1, \dots, m)$$

hold along the arcs g_1 and g_2 respectively. A simple continuous curve which

lies in R_1 or R_2 or both and which is composed of a finite succession of regular arcs of class C^2 along which the Euler equations hold will be termed a *broken extremal*.

If $h^i(t)$ is any function of t defined for t near $t=c$ on g , we shall represent the left and right limits of $h^i(t)$ at $t=c$, provided they exist, by h^{i-} and h^{i+} respectively.

If the point $t=c$ on g is denoted by P , by points near P on the negative (positive) side of M we shall mean points lying in R_1 (R_2) near P .

2. The primary incidence relations. It is easy to prove that a necessary condition that g afford a weak minimum to J relative to neighboring admissible curves of class D^1 is that the directions of g at $t=c$ satisfy the following conditions

$$(2.1) \quad [F^1_{r^i}(\gamma^-, \dot{\gamma}^-) - F^2_{r^i}(\gamma^+, \dot{\gamma}^+)] z_h^i(0) = 0, \\ (i = 1, \dots, m; h = 1, \dots, n)$$

where the subscript h indicates differentiation with respect to α^h . From now on we shall assume that the directions of g at $t=c$ satisfy (2.1).

Consider the condition

$$(2.2) \quad [F^1_{r^i}(z, r^-) - F^2_{r^i}(z, r^+)] dz^i = 0, \quad (i = 1, \dots, m)$$

where (z) is given by (1.2), where (r^-) and (r^+) denote directions near $(\dot{\gamma}^-)$ and $(\dot{\gamma}^+)$ respectively, where the differentials dz^i are to be expressed in terms of the differentials $d\alpha^h$ using (1.2), and where (2.2) is to be regarded as an identity in these differentials for (α) near $(\alpha) = (0)$.

For a point (z) on M near $(\alpha) = (0)$, the condition (2.2) may be written in the form

$$(2.3)' \quad \{F^1_{r^i}[z(\alpha), r^-] - F^2_{r^i}[z(\alpha), r^+]\} z_h^i(\alpha) = 0, \\ (i = 1, \dots, m; h = 1, \dots, n)$$

where the subscript h indicates differentiation with respect to α^h . The conditions (2.3)' will be termed the *primary incidence relations* at a point (z) on the deflecting manifold M .

A broken extremal which is composed of two successive extremal arcs lying in R_1 and R_2 respectively will be termed an *extremaloid* if its directions at the corner (z) on M satisfy the primary incidence relations (2.3)'. We note that g is an extremaloid which satisfies the primary incidence relations with (α) therein equal to (0) .

If to the conditions (2.3)' we adjoin the condition

$$(2.3)'' \quad r^{i+} r^{i-} = 1, \quad (i = 1, \dots, m)$$

the m conditions (2.3), considered as equations in the variables (r^+) , have a unique solution $r^{i+} = r^{i+}(\alpha, r^-)$, where r^{i+} are functions of class C^2 for (α) near (0) and (r^-) near $(\dot{\gamma}^-)$. That such a solution exists follows from the fact that g satisfies the primary incidence relations and from the fact that for $(\alpha) = (0)$ the functional determinant of the left members of the system (2.3) with respect to the variables (r^+) is not zero. To prove the latter fact we use the method of Bliss (see Bliss 2, p. 447) to show that the m -square functional determinant

$$(2.4) \quad \left| \begin{matrix} z_h^i F_{r^i r^j}^{2+} \\ \dot{\gamma}^{j+} \end{matrix} \right|^{(\alpha)=(0)} \quad (i, j = 1, \dots, m; h = 1, \dots, n)$$

is equal to $\pm B \cdot F_1$, where

$$B = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & \dot{\gamma}^1 & \dots & \dot{\gamma}^m \\ 0 & z_1^1 & \dots & z_1^m \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & z_n^1 & \dots & z_n^m \end{vmatrix}^+, \quad F_1 = \begin{vmatrix} 0 & \dot{\gamma}^j \\ \dot{\gamma}^i & F_{r^i r^j}^{2+} \end{vmatrix}^+, \quad (\alpha) = (0).$$

The determinant B is not zero since the regular manifold M is not tangent to g_2 at $t = c$. That F_1 is not zero is a consequence of the positive regularity hypothesis (1.4). See Morse 1, p. 112.

We state the following theorems.

THEOREM 2.1. *Given a point (z) on M near $(\alpha) = (0)$ and at this point a direction (r^-) near $(\dot{\gamma}^-)$. There is a unique extremal on the positive side of M which issues from the point (z) with a direction (r^+) determined by the primary incidence relations, and along which the parameter is the arc length.*

THEOREM 2.2. *An n -parameter family of extremals defined for $t \leq c$ and intersecting M for $t = c$ determines an n -parameter family of extremals defined for $t \geq c$ which with the respective extremals of the given family satisfy the primary incidence relations at $t = c$, and along which the parameter is the arc length.*

Such a family defined for $t \geq c$ is called the *continuation* of the given family defined for $t \leq c$. The two families of extremals form a *family of extremaloids*.

Similarly a family of extremals may be defined in terms of a family of extremals which is defined for t on g_2 and intersects M for $t = c$ and its continuation family for t on g_1 .

3. Conjugate points. Let t^0 be a value of t , $t' \leq t^0 < c$, and let $(\dot{\gamma}^0)$ denote the direction cosines of g_1 at the point for which $t = t^0$. Let K denote the unit $(m-1)$ -sphere with center at the origin. Let $(\beta^1, \dots, \beta^{m-1})$ be the parameters in a regular representation of K in the neighborhood of the point $(\dot{\gamma}^0)$, with $(\beta) = (0)$ corresponding to $(\dot{\gamma}^0)$. The family of extremals issuing from t^0 with directions determined by (β) can be represented in the form

$$(3.1) \quad x^i = \phi^i(\tau, \beta^1, \dots, \beta^n) = \phi^i(\tau, \beta), \\ (i = 1, \dots, m; n = m - 1)$$

where τ is the arc length and $(\beta) = (0)$ gives g_1 . The functions ϕ^i and ϕ_{τ^i} are of class C^2 for (β) near (0) and τ on g_1 . The zeros, $\tau \neq t^0$, of the jacobian

$$(3.2) \quad D(\tau, t^0) = \frac{D(\phi^1, \dots, \phi^m)}{D(\tau, \beta^1, \dots, \beta^n)}, \quad (\beta) = (0),$$

of the family (3.1) are termed the *conjugate points on g_1* of the point t^0 of g_1 . The order of vanishing of $D(\tau, t^0)$ at a conjugate point τ of t^0 is termed the *order* of that conjugate point.

Since g_1 is not tangent to M at $t = c$, the equations

$$\phi^i(\tau, \beta) = z^i(\alpha) \quad (i = 1, \dots, m)$$

have a solution of the form

$$(3.3) \quad \tau = \tau(\beta), \\ \alpha^h = \alpha^h(\beta), \quad (h = 1, \dots, n = m - 1)$$

where $\tau(\beta)$ and $\alpha^h(\beta)$ are functions of class C^2 for (β) near (0) , and where $\tau(0) = c$, $\alpha^h(0) = 0$. Geometrically this means that the extremals of the family (3.1) intersect M for (β) near (0) in the space (x) .

In order to define the conjugate points on g_2 of the point t^0 of g_1 it is convenient to represent the family (3.1) in the form

$$(3.4) \quad x^i = \phi^i(\tau, \beta) = \psi^i(t, \beta) \quad (i = 1, \dots, m)$$

where

$$t = \frac{t^0[\tau(\beta) - \tau] + c[\tau - t^0]}{\tau(\beta) - t^0},$$

and

$$\psi^i(t^0, \beta) \equiv \gamma^i(t^0), \\ \psi^i(c, \beta) \equiv z^i[\alpha(\beta)].$$

Such a change of parameter is admissible, and we note that $(\beta) = (0)$ gives g_1 for the family (3.4) and that along g_1 the parameter t is the arc length. Moreover the jacobian

$$(3.5) \quad D_1(t, t^0) = \frac{D(\psi^1, \dots, \psi^m)}{D(t, \beta^1, \dots, \beta^n)}, \quad (\beta) = (0),$$

of the family (3.4) vanishes if and only if the jacobian (3.2) vanishes, and it vanishes to the same order. Thus the conjugate points on g_1 of the point t^0 of g_1 are defined by the zeros, $t \neq t^0$, of $D_1(t, t^0)$, and their orders, by the order of vanishing of $D_1(t, t^0)$ at the respective points.

From Theorem 2.2 it follows that there exists a family of extremals on the positive side of M which issue from the points (z) on M near $(\alpha) = (0)$, with directions r^{i+} determined in (2.3) by the directions r^{i-} of the family (3.4), and along which the parameter is the arc length. These extremals represent the unique continuations of the respective extremals of the family (3.4), and will be represented in the form

$$(3.6) \quad x^i = \psi^i(t, \beta), \quad (t \text{ on } g_2)$$

where t is the arc length, where $(\beta) = (0)$ gives g_2 , and where

$$\psi^i(c, \beta) \equiv z^i[\alpha(\beta)] \quad (i = 1, \dots, m).$$

The functions ψ^i and $\psi_{,i}$ are of class C^2 for t on g_2 and (β) near (0) . The zeros, $t \neq c$, of the jacobian

$$(3.7) \quad D_2(t, t^0) = \frac{D(\psi^1, \dots, \psi^m)}{D(t, \beta^1, \dots, \beta^n)}, \quad (\beta) = (0),$$

of the family (3.6) are termed the *conjugate points on g_2* of the point t^0 of g_1 . The order of vanishing of $D_2(t, t^0)$ at a conjugate point t of t^0 is the *order* of that conjugate point.

We shall find it convenient to refer to the family of extremaloids

$$(3.8) \quad x^i = \psi^i(t, \beta) \quad (t \text{ on } g)$$

which is defined by (3.4) and (3.6) for t on g_1 and g_2 respectively.

Conjugate point determinant. We set

$$(3.9) \quad \begin{aligned} D_g(t, t^0) &= D_1(t, t^0), & (t' \leq t \leq c) \\ D_g(t, t^0) &= D_2(t, t^0), & (c < t \leq t'') \end{aligned}$$

understanding that $(\beta) = (0)$ therein, and term $D_g(t, t^0)$ the *conjugate point determinant*. The zeros, $t \neq t^0$, of $D_g(t, t^0)$ define the conjugate points on g of the point t^0 of g_1 . The conjugate points of t^0 and their orders are independent of admissible changes of parameter.

For any point $t^0 \neq c$ on g_2 the conjugate point determinant $D_g(t, t^0)$ is defined in a similar fashion, using the family of extremaloids issuing from the point t^0 with directions near the direction of g_2 at t^0 . In so doing it is understood that the corner point $t = c$ is considered as a point of g_1 .

Finally if $t^0 = c$, the conjugate points on g of the point $t = c$ are the zeros, $t \neq c$, of the conjugate point determinant $D_g(t, c)$ defined in terms of the family of extremaloids with a corner at the point $t = c$ on g , with directions r^{i-} determined as in (3.1) by the parameters (β) in a regular representation of K in the neighborhood of the point $(\dot{\gamma}^-)$, $(\beta) = (0)$ corresponding to $(\dot{\gamma}^-)$, and along which the parameter t is the arc length. The order of vanishing of $D_g(t, c)$ at a conjugate point is the order of that conjugate point.

4. The second variation. Let

$$\alpha^h = \alpha^h(e), \quad \alpha^h(0) = 0, \quad (h = 1, \dots, n = m - 1)$$

be a set of n functions of class C^2 for e near 0. Let

$$(4.1) \quad x^i = x^i(t, e) \quad (i = 1, \dots, m)$$

be a 1-parameter family of admissible curves for which the functions $x^i(t, e)$ are of class C^2 for t on g_1 and e near 0 and for t on g_2 and e near 0 respectively, which contains g for $e = 0$, and which satisfies the identities

$$(4.2) \quad x^i(c, e) \equiv z^i[\alpha(e)].$$

For each value of e near 0, the integral J evaluated along the admissible curve determined by e is a function $J(e)$ of class C^2 . We obtain a formula for the second variation in which we set

$$2\Omega^\kappa(\eta, \dot{\eta}) = F^\kappa_{r^i r^j} \eta^i \dot{\eta}^j + 2F^\kappa_{x^i r^j} \eta^i \dot{\eta}^j + F^\kappa_{x^i x^j} \eta^i \dot{\eta}^j, \\ (\kappa = 1, 2; i, j = 1, \dots, m)$$

where the arguments of the partial derivatives of F^κ are $(x, r) = (\gamma, \dot{\gamma})$ on g_κ .

Since g satisfies the primary incidence relations at $t = c$, the second variation takes the form

$$J''(0) = b_{hk} \omega^h \omega^k + \int_{t'}^0 2\Omega^1(\eta, \dot{\eta}) dt + \int_0^{t''} 2\Omega^2(\eta, \dot{\eta}) dt,$$

where

$$(4.3) \quad b_{hk} = [F^1_{r^i}(\gamma^-, \dot{\gamma}^-) - F^2_{r^i}(\gamma^+, \dot{\gamma}^+)] z^i_{hk}(0), \\ (i = 1, \dots, m; h, k = 1, \dots, n)$$

and where η^i and the n constants ω^h are respectively the variations $x_\sigma^i(t, 0)$ and $\alpha_\sigma^h(0)$ and satisfy the secondary end conditions

$$(4.4) \quad \eta^i(t') = 0, \quad \eta^i(t'') = 0, \quad (i = 1, \dots, m)$$

and the secondary corner conditions

$$(4.5a) \quad \eta^{i-} = z^i_h(0) \omega^h, \quad (h = 1, \dots, n)$$

$$(4.5b) \quad \eta^{i+} = z^i_h(0) \omega^h.$$

5. Solutions of the Jacobi equations. The Jacobi equations for t on g_1 and g_2 are

$$(5.1) \quad \frac{d}{dt} \Omega^\kappa \dot{\eta}^i - \Omega^\kappa \eta^i = 0, \quad (i = 1, \dots, m)$$

where $\kappa = 1$ and $\kappa = 2$ respectively. Throughout this section we shall assume that κ is fixed; that is, κ is either 1 or 2, not both.

It is well-known that the Jacobi equations for t on g_κ are satisfied identically by *tangential solutions* of the form

$$\rho(t) \dot{\eta}^i(t) \quad (i = 1, \dots, m)$$

where ρ is an arbitrary function of t of class C^2 for t on g_κ . If, for $t = t^*$, a solution of the Jacobi equations $\eta^i(t)$ satisfies the relation

$$\eta^i(t^*) - \rho(t^*) \dot{\eta}^i(t^*) = 0, \quad (i = 1, \dots, m)$$

we shall say that $\eta^i(t)$ vanishes modulo a tangential solution at $t = t^*$. If a solution of the Jacobi equations for t on g_κ is determined except for the possible addition of a tangential solution, we shall say it is *determined modulo a tangential solution*, or more briefly, *mod T* . We seek conditions by which solutions of the Jacobi equations for t on g_κ are determined mod T .

Since the determinant of the coefficients of $\ddot{\eta}^i$ in (5.1) is zero, in order to obtain solutions of the Jacobi equations for t on g_κ we consider the auxiliary differential equations

$$(5.2) \quad \begin{aligned} \frac{d}{dt} \Omega^\kappa \dot{\eta}^i - \Omega^\kappa \eta^i &= 0, & (i, j = 1, \dots, m) \\ \frac{d^2}{dt^2} (\dot{\gamma}^j \eta^j) &= 0, \end{aligned}$$

for t on g_κ . Cf. Bliss 3, p. 199 and Graves 1, p. 17. To solve the $m+1$ equations (5.2) we introduce the system

$$(5.3) \quad \begin{aligned} \frac{d}{dt} \Omega^\kappa \dot{\eta}^i - \Omega^\kappa \eta^i + \lambda \dot{\gamma}^i &= 0, \\ \frac{d^2}{dt^2} (\dot{\gamma}^j \eta^j) &= 0, \end{aligned}$$

where λ is an unknown function of t of class C^2 for t on g_κ . Using the method of Morse 1, p. 124, it is easy to prove that $\lambda \equiv 0$ in solutions of (5.3). Thus (5.3) may be regarded as identical with (5.2). Hence the $\ddot{\eta}^i$ in (5.2) can be expressed as linear homogeneous functions of the variables $(\eta, \dot{\eta})$ with coefficients which are of class C^1 in t for t on g_κ .

The most general tangential solution of the auxiliary differential equations (5.2) is of the form

$$(a + bt)\dot{\gamma}^i(t), \quad (i = 1, \dots, m)$$

where a and b are constants.

We shall prove the following lemma.

LEMMA 5.1. *Any solution of the Jacobi equations for t on g_κ may be written as a solution of the auxiliary differential equations for t on g_κ plus a tangential solution of the Jacobi equations for t on g_κ .*

Let (η) be any solution of the Jacobi equations for t on g_κ . Consider the difference

$$y^i(t) = \eta^i(t) - \rho(t)\dot{\gamma}^i(t) \quad (i = 1, \dots, m)$$

where $\rho(t)$ is a function of t of class C^2 for t on g_κ . The difference $y^i(t)$ is a solution of the Jacobi equations for t on g_κ . If we choose $\rho(t)$ so that

$$\frac{d^2}{dt^2}(\dot{\gamma}^j y^j) = 0, \quad (j = 1, \dots, m)$$

then we have

$$(5.4) \quad \ddot{\rho} = \frac{d^2}{dt^2}(\dot{\gamma}^j \eta^j)$$

and $y^i(t)$ is a solution of (5.2), as was to be proved.

We shall prove the following lemma.

LEMMA 5.2. *A solution of the Jacobi equations for t on g_κ which vanishes with its derivative at a point, is identically equal to a tangential solution of the Jacobi equations for t on g_κ .*

Let (η) be any solution of the Jacobi equations for t on g_κ which vanishes with its derivative at a point $t = t^*$. Consider the difference

$$(5.5) \quad w^i(t) = \eta^i(t) - \rho(t)\dot{\gamma}^i(t),$$

where $\rho(t)$ is a function of class C^2 for t on g_κ which satisfies (5.4) and where $\rho(t^*) = \dot{\rho}(t^*) = 0$. Then $w^i(t)$ is a solution of (5.2) which vanishes with its derivative at $t = t^*$. Hence $w^i(t) \equiv 0$. It follows that $\eta^i(t)$ is identically equal to a tangential solution of the Jacobi equations for t on g_κ .

6. The secondary incidence relations. The secondary problem is non-parametric in the space of the variables $(\eta^1, \dots, \eta^m, t)$. The n -plane

$$N \quad \eta^i = z_h^i(0)\omega^h, \quad t = c \quad (i = 1, \dots, m; h = 1, \dots, n = m - 1)$$

is the analogue of the deflecting manifold M and will be referred to as the *deflecting plane* N . The deflecting plane N is regular by virtue of the fact that the rank of the matrix $\|z_h^i(0)\|$ is n . The only tangential solutions of

the Jacobi equations or of the auxiliary differential equations for which $t = c$ gives a point on N , are those tangential solutions which vanish at $t = c$.

In the space (η, t) , a solution of the Jacobi equations which is of class C^2 for t on g_1 and g_2 respectively, and which for $t = c$ has a corner on N is the analogue of a broken extremal in the space (x) with fixed end points and a single corner on the deflecting manifold M .

We shall next define the secondary incidence relations. To that end let (u) be an arbitrary set of $n = m - 1$ constants and e a parameter neighboring $e = 0$. Consider the 1-parameter family of extremaloids

$$(6.1) \quad x^i = x^i(t, e) \quad (i = 1, \dots, m)$$

determined by setting $\beta^h = eu^h$ in the family of extremaloids (3.8). The family (6.1) satisfies the following identities in e :

$$(6.2) \quad x^i(c, e) \equiv z^i[\alpha(eu)].$$

The variations $x_e^i(t, 0)$ of the family (6.1) will be denoted by $\eta^i(t)$. Differentiating the identities (6.2) with respect to e and setting $e = 0$ yields

$$(6.3) \quad \eta^{i\pm} = z_h^i(0) \frac{d\alpha^h}{de} \quad (h, k = 1, \dots, n)$$

where

$$\frac{d\alpha^h}{de} = \frac{\partial \alpha^h}{\partial \beta^k} u^k.$$

For the family (6.1) the primary incidence relations (2.3)' reduce to a set of n identities in e . Upon differentiating these identities with respect to e and setting $e = 0$, we obtain

$$(6.4) \quad b_{hk} \frac{d\alpha^k}{de} + z_h^i(0) \xi_i^- = 0, \quad (i = 1, \dots, m; h, k = 1, \dots, n)$$

where b_{hk} is given by (4.3) and where

$$(6.5) \quad \xi_i^- = \Omega^1 \eta^i(\eta^-, \dot{\eta}^-), \quad \xi_i^+ = \Omega^2 \eta^i(\eta^+, \dot{\eta}^+).$$

Setting $d\alpha^k/de = \omega^k$, (6.3) becomes

$$(6.6) \quad \eta^{i\pm} = z_h^i(0) \omega^h,$$

and (6.4) becomes

$$(6.7) \quad b_{hk} \omega^k + z_h^i(0) \xi_i^- = 0.$$

We term (6.7) subject to (6.6) and (6.5), the *secondary incidence relations*. It is understood that the independent variables in (6.7) are (ω) , (η^-) and $(\dot{\eta}^-)$.

If $\eta^i(t)$ is a solution of the Jacobi equations with a corner on N for $t = c$, and if its slopes $\dot{\eta}^{i-}$ and $\dot{\eta}^{i+}$ with the set (ω) determined by (6.6)

satisfy the secondary incidence relations (6.7), then $\eta^i(t)$ will be said to satisfy the secondary incidence relations.

In order to solve the secondary incidence relations for the variables $(\dot{\eta}^+)$ in terms of the remaining variables, we adjoin the condition

$$\dot{\eta}^j \dot{\eta}^j]_{+}^{-} = 0 \quad (j = 1, \dots, m)$$

to the relations (6.7). The m conditions

$$(6.8) \quad \begin{aligned} b_{hk} \omega^k + z_h^i(0) \xi_i]_{+}^{-} &= 0, & (i, j = 1, \dots, m; h, k = 1, \dots, n) \\ \dot{\eta}^j \dot{\eta}^j]_{+}^{-} &= 0, \end{aligned}$$

are called the *restricted* secondary incidence relations. Since the problem is positive regular along g_2 , and since the regular manifold M is not tangent to g_2 at $t = c$, the restricted secondary incidence relations (6.8), considered as equations in the variables $(\dot{\eta}^+)$, have a unique solution $(\dot{\eta}^+)$ expressible in terms of the remaining variables.

We have the following lemma.

LEMMA 6.1. *Corresponding to any solution of the auxiliary differential equations for t on g_1 which for $t = c$ intersects the deflecting plane N at the point (ω) , there exists a unique solution of the auxiliary differential equations for t on g_2 which for $t = c$ gives the point (ω) on the deflecting plane N and at (ω) has a slope uniquely determined by the restricted secondary incidence relations.*

Such a solution is called the *continuation* for t on g_2 of the given solution for t on g_1 .

Returning to the problem of expressing the variables $(\dot{\eta}^+)$ of the secondary incidence relations (6.7) in terms of the remaining variables, we state the following lemma.

LEMMA 6.2. *Corresponding to sets $(\dot{\eta}^-)$ and (ω) , there is a set $(\dot{\eta}^+)$ determined except for the possible addition of a set of the form $(k\dot{\eta}^+)$ where k is a constant, by the secondary incidence relations (6.7).*

The proof of the lemma is based on the following statements.

(α). If the variables $(\dot{\eta}^+)$ with sets $(\dot{\eta}^-)$ and (ω) satisfy the secondary incidence relations (6.7), then the set $(\dot{\eta}^+ + k\dot{\eta}^+)$, where k is a constant, with the same sets $(\dot{\eta}^-)$ and (ω) satisfies the secondary incidence relations (6.7).

(β). Any two sets $(\dot{\eta}^+)$ and $(\dot{\eta}^*)$ which with sets $(\dot{\eta}^-)$ and (ω) satisfy the secondary incidence relations (6.7) differ by a set of the form $(k\dot{\eta}^+)$, where k is a constant.

Statements (α) and (β) may be readily verified by direct substitution.

The following theorem is a consequence of Lemma 6.2 and the fact that the only tangential solutions of the Jacobi equations which for $t = c$ define a point on the deflecting plane N are those which vanish at $t = c$.

THEOREM 6.1. *Corresponding to any solution (η) of the Jacobi equations for t on g_1 which for $t = c$ defines a point (ω) on the deflecting plane N , there exists a solution of the Jacobi equations for t on g_2 which for $t = c$ gives the point (ω) on the deflecting plane N and which at $t = c$ has a slope determined except for a constant multiple of $\dot{\gamma}^4$ by the secondary incidence relations (6.7). This solution is unique modulo a tangential solution of the Jacobi equations for t on g_2 which vanishes at $t = c$.*

Such a solution of the Jacobi equations is termed a *continuation* for t on g_2 of the given solution for t on g_1 .

COROLLARY 6.1. *A tangential solution of the Jacobi equations for t on g_1 has continuations for t on g_2 , if and only if it vanishes at $t = c$. Moreover its continuations for t on g_2 are arbitrary tangential solutions which vanish at $t = c$.*

We state the following theorem.

THEOREM 6.2. *The continuations for t on g_2 , if they exist, of any two mutually conjugate solutions of the Jacobi equations for t on g_1 are mutually conjugate in the sense of von Escherich.*

Let η^i and $\bar{\eta}^i$ be any two solutions of the Jacobi equations for t on g_1 which are mutually conjugate in the sense of von Escherich; that is, suppose the identity

$$(6.9) \quad \eta^i \Omega^1 \dot{\eta}^i - \bar{\eta}^i \Omega^1 \dot{\bar{\eta}}^i \equiv 0 \quad (i = 1, \dots, m)$$

holds for t on g_1 . (See Bolza 1, p. 626). We assume that the continuations of (η) and ($\bar{\eta}$) do exist. Making use of the fact that the given solutions (η) and ($\bar{\eta}$) and their respective continuations satisfy the secondary incidence relations, it is easy to prove that the continuations of (η) and ($\bar{\eta}$) are mutually conjugate in the sense of von Escherich. (Cf. Morse 1, p. 52.)

7. The determinant $\Delta(t, t^0)$. In this section we shall assume, unless otherwise specified, that a point t^0 on g_1 means a point t^0 such that $t' \leq t^0 < c$. Recall the conjugate point determinant $D_\theta(t, t^0)$ defined in (3.9). Let the $(p+1)$ -st column of $D_\theta(t, t^0)$ be represented in the form

$$(7.1) \quad \eta_p^i(t), \quad (i = 1, \dots, m; p = 1, \dots, n)$$

and let the first column be multiplied by $(t - t^0)(t - c)$ so that it will be a tangential solution of the Jacobi equations which vanishes at $t = t^0$ and is continuable at $t = c$. The determinant $\Delta(t, t^0)$ is defined as follows:

$$\Delta(t, t^0) = |(t - t^0)(t - c)\dot{\gamma}^i(t) \quad \eta_p^i(t)|.$$

For $t \neq t^0$ and $t \neq c$, the determinant $\Delta(t, t^0)$ vanishes if and only if the conjugate point determinant $D_\theta(t, t^0)$ vanishes, and to the same order. Hence the zeros, $t \neq t^0$ and $t \neq c$, of $\Delta(t, t^0)$ define the conjugate points on g of the point t^0 of g_1 , and the order of vanishing of $\Delta(t, t^0)$ at a conjugate point defines the order of that conjugate point. For $t = c$, $\Delta(t, t^0)$ always vanishes. The point c of g will be conjugate to the point t^0 of g_1 if and only if the order of vanishing of $\Delta(c, t^0)$ is greater than 1, and the order of c as a conjugate point of t^0 will be 1 less than the order of vanishing of $\Delta(c, t^0)$.

The variation $\eta_p^i(t)$ representing the $(p + 1)$ -st column of $\Delta(t, t^0)$ is a solution of the Jacobi equations determined by g . The variation $\eta_p^i(t)$ is, moreover, precisely the variation $x_e^i(t, 0)$ of the family (6.1) when $u^p = 1$ and the other $n - 1$ u 's are null. Since the secondary incidence relations are linear in all the variables, we have the following theorem.

THEOREM 7.1. *The combination $w^p \eta_p^i(t)$ of the last n columns of $\Delta(t, t^0)$ is a solution of the Jacobi equations which satisfies the secondary incidence relations (6.7), provided (ω) therein is taken as*

$$\omega^k = \frac{\partial x^k}{\partial \beta^p} w^p \quad (k, p = 1, \dots, n).$$

We shall prove the following theorem.

THEOREM 7.2. *The m columns of the determinant $\Delta(t, t^0)$ represent m linearly independent solutions of the Jacobi equations for t on g , and the last n columns represent solutions which are linearly independent of tangential solutions.*

That the columns of $\Delta(t, t^0)$ are linearly independent for t on g_1 follows from the fact that $\Delta(t, t^0)$ does not vanish identically for t sufficiently near t^0 . Suppose first then that the last n columns are linearly dependent upon a tangential solution for t on g_1 ; that is, suppose n constants (c) , not all (0) , exist so that

$$(7.2) \quad c_p \eta_p^i(t) \equiv \rho(t) \dot{\gamma}^i(t) \quad (p = 1, \dots, n)$$

where ρ is a function of t of class C^2 for t on g_1 . The function ρ cannot be identically zero for t on g_1 , for that would imply the linear dependence of the last n columns of $\Delta(t, t^0)$ for t on g_1 . But since $(c) \neq (0)$ and $\rho \neq 0$,

the identity (7.2) implies that $\Delta(t, t^0)$ vanishes identically for t on g_1 . From this contradiction we infer that the last n columns of $\Delta(t, t^0)$ are linearly independent of tangential solutions for t on g_1 .

It remains to prove that the theorem is true for t on g_2 . Next suppose that the m solutions of the Jacobi equations represented by the columns of $\Delta(t, t^0)$ are linearly dependent for t on g_2 ; that is, suppose that constants $(-d, c_1, \dots, c_n)$, not all zero, exist so that the identity

$$c_p \eta_p^i(t) \equiv d(t - t^0)(t - c) \dot{\gamma}^i(t) \quad (i = 1, \dots, m; p = 1, \dots, n)$$

holds for t on g_2 . The constants (c) cannot all be null, for $c_p = 0$ for each p would imply that $d = 0$. By Theorem 7.1 the solution $c_p \eta_p^i(t)$ for t on g_2 is a continuation of $c_p \eta_p^i(t)$ for t on g_1 . On the other hand since $c_p \eta_p^i(t)$ for t on g_2 is identically equal to a tangential solution which vanishes at $t = c$, it must be a continuation of a tangential solution which vanishes at $t = c$. Hence for t on g_1 we have $c_p \eta_p^i(t) + \rho(t) \dot{\gamma}^i(t) \equiv 0$, where $\rho(t)$ is a function of class C^2 for t on g_1 and where $\rho(c) = 0$. Now since $(c) \neq (0)$, this implies the linear dependence of the last n columns of $\Delta(t, t^0)$ on a tangential solution for t on g_1 . From this contradiction we infer the linear independence of the m columns of $\Delta(t, t^0)$ for t on g_2 .

The proof that the last n columns of $\Delta(t, t^0)$ are linearly independent of tangential solutions is similar and will be omitted.

That the following theorem is true for t on g_1 may be verified by substitution in (6.9). That it is true for t on g_2 follows from Theorem 6.2.

THEOREM 7.3. *The columns of $\Delta(t, t^0)$ represent mutually conjugate solutions of the Jacobi equations.*

We shall prove the following theorem.

THEOREM 7.4. *A necessary and sufficient condition that a point $t = t^*$ on $c < t^* \leq t''$ be conjugate to a point $t = t^0$ on $t' \leq t^0 < c$ is that there exist a solution of the Jacobi equations which vanishes at $t = t^0$ and $t = t^*$, which satisfies the secondary incidence relations at $t = c$, and which is not given by a tangential solution of the Jacobi equations.*

If the point $t = t^*$ is conjugate to the point $t = t^0$, then $\Delta(t, t^0)$ vanishes for $t = t^*$. There exists then a proper linear combinations (w) of the columns of $\Delta(t, t^0)$ which vanishes. Let (d, c_1, \dots, c_n) denote the constants in this linear combination. First, I say that c_p is not zero for each p , for $(c) = (0)$ would imply $d = 0$ which is impossible. The solution (w) defines a solution of the Jacobi equations which vanishes at $t = t^0$ and $t = t^*$, which satisfies the secondary incidence relations at $t = c$, and which is not given by a tangential solution.

Conversely let $\bar{\eta}^i(t)$ be a solution of the Jacobi equations which vanishes at $t = t^0$ and $t = t^*$, which satisfies the secondary incidence relations at $t = c$, and which is not given by a tangential solution. Consider the difference

$$(7.3) \quad \bar{\eta}^i(t) - d(t - t^0)(t - c)\dot{\gamma}^i(t) - c_p\eta_p^i(t),$$

where $\eta_p^i(t)$ is given by (7.1) and (d, c_1, \dots, c_n) are constants. The difference (7.3) is a solution of the Jacobi equations which vanishes at $t = t^0$. Moreover, by virtue of the fact that the determinant

$$|\dot{\gamma}^i(t^0) \quad \dot{\eta}_p^i(t^0)|$$

is not zero, the constants in (7.3) may be chosen so that the derivative of (7.3) also vanishes at $t = t^0$. Consequently for t on g_1 , we have

$$(7.4) \quad \bar{\eta}^i(t) - c_p\eta_p^i(t) \equiv 0, \quad (c) \neq (0),$$

modulo a tangential solution which vanishes at $t = t^0$ and $t = c$. The solution represented by the left and right members of (7.4) must have the same continuations for t on g_2 . A continuation of the left member of (7.4) is $\bar{\eta}^i(t) - c_p\eta_p^i(t)$, and all the continuations of the right member are tangential. For t on g_2 , then, we have $\bar{\eta}^i(t) - c_p\eta_p^i(t) \equiv 0$, modulo a tangential solution which vanishes at $t = c$.

Since $\bar{\eta}^i(t^*) = 0$, it follows that for some constant k , $c_p\eta_p^i(t^*) - k\dot{\gamma}^i(t^*) = 0$. But $(c) \neq (0)$, so that $\Delta(t, t^0)$ must vanish at $t = t^*$, and the point $t = t^*$ is then conjugate to the point $t = t^0$. The proof of Theorem 7.4 is complete.

A necessary and sufficient condition that a point $t = t^*$ on $t' \leq t^* \leq c$ be conjugate to a point $t = t^0$ on $t' \leq t^0 < c$ is that there exist a solution of the Jacobi equations which vanishes at $t = t^0$ and $t = t^*$, which satisfies the secondary incidence relations at $t = c$, and which is not given by a tangential solution of the Jacobi equations.

Consider the m -square determinant

$$\theta(t) = |(t - t^0)(t - c)\dot{\gamma}^i(t) \quad \mu_p^i(t)| \\ (i = 1, \dots, m; p = 1, \dots, n)$$

in which the last n columns are solutions of the Jacobi equations which vanish at the point $t = t^0$ on g_1 , which satisfy the secondary incidence relations at $t = c$, and which are linearly independent of tangential solutions.

LEMMA 7.1. *The order of vanishing of $\theta(t)$ at any point $t = b$ on g is equal to the nullity ν of $\theta(b)$.*

The proof of this lemma follows the method of Morse in (2), but slight

modifications are necessary since we are using solutions of the Jacobi equations in place of solutions of his "restricted Jacobi equations."

Let b be any value of t on g except t^0 and c . Let r be the rank of $\theta(b)$. Then $r > 0$, and $\nu = m - r$. We suppose that the last n columns of $\theta(b)$ have been reordered so that the rank of the first r columns is r . We also suppose that the rank of the last ν columns is zero; for if it were not zero, it could be made zero by adding suitably chosen linear combinations of the first r columns to the remaining columns. Understanding that this has been done, let

$$u_h^i(t), \quad v_k^i(t) \quad (i = 1, \dots, m; h = 1, \dots, r; k = 1, \dots, \nu)$$

represent the first r and last ν columns respectively of $\theta(t)$.

Applying the integral form of the law of the mean to the elements in the last ν columns of $\theta(t)$ yields

$$(7.5) \quad \theta(t) = (t - b)^\nu B(t)$$

where the function $B(t)$ is continuous in t for t on g_1 and for t on g_2 respectively, and where

$$B(b) = |u_h^i(b) \quad v_k^i(b)|.$$

The lemma will follow from (7.5) if $B(b) \neq 0$.

Suppose that $B(b) = 0$. There will exist then a proper linear combination (w) of the columns of $B(b)$ with coefficients $(c_1, \dots, c_r, -d_1, \dots, -d_\nu)$ such that $(w) = (0)$. Moreover the constants d_k cannot all be zero. For $d_k = 0$ for each k would imply that $c_h u_h^i(b) = 0$ for each i , and the rank of $\theta(b)$ would be less than r . We set

$$u^i(t) = c_h u_h^i t, \quad v^i(t) = d_k v_k^i(t) \\ (i = 1, \dots, m; h = 1, \dots, r; k = 1, \dots, \nu).$$

Hence

$$(7.6) \quad u^i(b) = \dot{v}^i(b), \quad v^i(b) = 0.$$

We note that $u^i(b)$ cannot be zero for each i . For that would imply $v^i(b) = \dot{v}^i(b) = 0$ for each i , and hence that (v) is a tangential solution which vanishes with its derivative at $t = b$. That this is impossible follows from the hypothesis that the last n columns of $\theta(t)$ are linearly independent of tangential solutions.

Making use of (7.6) and of the fact that $u^i(t)$ and $v^i(t)$ are mutually conjugate for t on g , we obtain

$$F^{\kappa, r, r'}[\gamma(b), \dot{\gamma}(b)] \dot{v}^i(b) \dot{v}^j(b) = 0,$$

where κ is 1 or 2 according as $t = b$ lies on g_1 or g_2 . It follows from (1.4) that

$$(7.7) \quad \dot{v}^i(b) = k \dot{\gamma}^i(b) \quad (k \neq 0).$$

But (7.6) and (7.7) imply that (v) is identically equal to a tangential solution which vanishes at $t = b$. This is impossible since the last n columns of $\theta(t)$ are linearly independent of tangential solutions. We conclude that $B(b)$ is not zero, and the lemma follows from (7.5) when $b \neq t^0$ and $b \neq c$.

The lemma is true when $b = t^0$ and $b = c$, but the proof is slightly different.

Since the determinant $\Delta(t, t^0)$ satisfies the conditions imposed on $\theta(t)$ in Lemma 7.1, we have the following theorem.

THEOREM 7.5. *The conjugate points on g of a point t^0 of g_1 are isolated and possess orders equal to the nullity ν of $\Delta(t, t^0)$ at the respective zeros $t \neq t^0$ and $t \neq c$ of $\Delta(t, t^0)$. The order of $t = c$ as a conjugate point of $t = t^0$ is $\nu - 1$.*

Since any solution of the Jacobi equations which vanishes at $t = t^0$ and satisfies the secondary incidence relations at $t = c$ can be written as a linear combination of the last n columns of $\Delta(t, t^0)$ and a tangential solution which vanishes at $t = t^0$ and $t = c$, we have the following theorem.

THEOREM 7.6. *If a point $t = t^*$ on $c < t^* \leq t''$ is conjugate to a point $t = t^0$ on $t' \leq t^0 < c$, the maximum number of solutions of the Jacobi equations which vanish at $t = t^0$ and $t = t^*$, which satisfy the secondary incidence relations at $t = c$, and which are linearly independent of tangential solutions, equals the order of the conjugate point.*

If a point $t = t^*$ on $t' \leq t^* \leq c$ is conjugate to a point $t = t^0$ on $t' \leq t^0 < c$, the maximum number of solutions of the Jacobi equations which vanish at $t = t^0$ and $t = t^*$, which satisfy the secondary incidence relations at $t = c$, and which are linearly independent of tangential solutions, equals the order of the conjugate point.

Throughout this section we have assumed that $t^0 \neq c$ is a point of g_1 . Corresponding theorems hold if t^0 is a point on g_2 for which $c < t^0 \leq t''$.

If $t^0 = c$, we define $\Delta(t, c)$ as the determinant obtained by multiplying the first column of the conjugate determinant $D_g(t, c)$ by $(t - c)$. For $t \neq c$, $\Delta(t, c)$ vanishes if and only if $D_g(t, c)$ vanishes and to the same order. The m columns of $\Delta(t, c)$ represent m linearly independent, mutually conjugate solutions of the Jacobi equations which vanish at $t = c$ and satisfy the secondary incidence relations with $(\omega) = (0)$ therein. Moreover the last n columns represent solutions which are linearly independent of tangential solutions of the Jacobi equations. A necessary and sufficient condition that a point $t = t^*$ on g be conjugate to the point $t = c$ is that there exist a solution of the Jacobi equations which vanishes at $t = c$ and $t = t^*$, which satisfies

the secondary incidence relations at $t = c$, and which is not given by a tangential solution of the Jacobi equations. Furthermore the conjugate points on g of the point $t = c$ are isolated and possess orders equal to the nullity of $\Delta(t, c)$ at the respective zeros $t \neq c$ of $\Delta(t, c)$. The order of vanishing of $\Delta(c, c)$ is equal to the nullity m of $\Delta(c, c)$. Finally, if a point $t = t^*$ of g is conjugate to the point $t = c$, the maximum number of solutions of the Jacobi equations which vanish at $t = c$ and $t = t^*$, which satisfy the secondary incidence relations at $t = c$, and which are linearly independent of tangential solutions, equals the order of the conjugate point.

8. The index theorem. Let $t = a_\sigma$, $\sigma = 1, \dots, \mu - 1, \mu + 1, \dots, \lambda$ be a set of values of t such that

$$(8.1) \quad a_0 < a_1 < \dots < a_{\mu-1} < c < a_{\mu+1} < \dots < a_\lambda < a_{\lambda+1}, \\ (a_0 = t', a_{\lambda+1} = t'')$$

and such that no one of the $\lambda + 1$ segments into which g is divided by the points of (8.1) contains a conjugate point of its initial end point.

Let M_σ be a regular $(m - 1)$ -manifold of class C^2 which intersects g at the point $t = a_\sigma$, but which is not tangent to g at that point. We suppose that M_σ is regularly represented neighboring the point $t = a_\sigma$ in the form

$$x^i = Z^{i\sigma}(\beta_\sigma^1, \dots, \beta_\sigma^n) \quad (i = 1, \dots, m; n = m - 1; \sigma \text{ not summed})$$

and that $(\beta_\sigma) = (0)$ determines the point $t = a_\sigma$ on g . Set $a_\mu = c$, and let M_μ be an alternative notation for the deflecting manifold M . The manifolds M_q , $q = 1, \dots, \lambda$, are termed a set of *intermediate manifolds*.

Let the points $t = a_0$ and $t = a_{\lambda+1}$ on g be denoted by A and B respectively. Let points P_q on the respective intermediate manifolds M_q be chosen so near to g that the successive points

$$(8.2) \quad A, P_1, \dots, P_\lambda, B$$

can be joined by extremal segments.

Let (v) be a set of λn variables, the q -th n of which are the parameters of the point P_q on M_q ; that is,

$$(v^1, \dots, v^{\lambda n}) = (\beta_1^1, \dots, \beta_1^n, \beta_2^1, \dots, \beta_2^n, \dots, \beta_{\mu-1}^1, \dots, \beta_{\mu-1}^n, \alpha^1, \dots, \alpha^n, \beta_{\mu+1}^1, \dots, \beta_{\mu+1}^n, \dots, \beta_\lambda^1, \dots, \beta_\lambda^n).$$

For (v) sufficiently near (0) , the points (8.2) are completely determined by the set (v) and the points A and B . The broken extremal $E(v)$ joining the points (8.2) will be expressed in the form

$$(8.3) \quad x^i = X^i(t, v), \quad (i = 1, \dots, m)$$

where X^i and X_{t^i} are functions of class C^2 in their arguments for (v) near

(0) and t on each of the $\lambda + 1$ components of $E(v)$, and where $(v) = (0)$ gives g . We assume that the parameter t has been chosen so that $t = t'$ and $t = t''$ give the points A and B respectively, so that the values $t = a_q$ ($q = 1, \dots, \lambda$) give the respective points P_q on M_q , and so that between each two successive vertices (8.2) the rate of change of t with respect to the arc length is constant.

The integral J considered along $E(v)$ is a function of class C^2 for (v) sufficiently near (0) , and will be denoted by $J(v)$.

THEOREM 8.1. *The function $J(v)$ has a critical point for $(v) = (0)$.*

To prove this theorem we consider the first partial derivatives of $J(v)$ with respect to the variables v^r , $r = 1, \dots, \lambda n$. Integrating by parts, setting $(v) = (0)$, and making use of the fact that g satisfies the primary incidence relations at $t = c$, we find that the first partial derivatives of $J(v)$ with respect to the variables v^r all vanish for $(v) = (0)$.

We set

$$Q(v) = J^0_{v^r v^s} v^r v^s, \quad (r, s = 1, \dots, \lambda n)$$

where the superscript 0 indicates evaluation for $(v) = (0)$, and term $Q(v)$ the *index form associated with the extremaloid g* . If \bar{r} is the rank of $Q(v)$, the number $\lambda n - \bar{r}$ is termed the *nullity* of $Q(v)$, and will be denoted by $N(Q)$. The *index* of $Q(v)$ is the number of negative characteristic roots belonging to $|J^0_{v^r v^s}|$, and will be denoted by $I(Q)$.

In order to obtain a representation of the index form $Q(v)$ in terms of the second variation we set up the family of broken extremals E determined by the points A and B and the set $(ev^1, \dots, ev^{\lambda n})$, where (v) is held fast and e is a variable near 0. The family of broken extremals E will be represented in the form

$$(8.4) \quad x^i = x^i(t, e) \quad (i = 1, \dots, m)$$

where the functions $x^i(t, e)$ are defined by referring to (8.3) and setting $X^i(t, ev) = x^i(t, e)$. The family (8.4) has the property that for e near 0,

$$\begin{aligned} x^i(c, e) &\equiv z^i(e\alpha^1, \dots, e\alpha^n), & (i = 1, \dots, m; n = m - 1) \\ x^i(a_\sigma, e) &\equiv Z^{i\sigma}(\beta_\sigma^1, \dots, \beta_\sigma^n) \\ &(\sigma = 1, \dots, \mu - 1, \mu + 1, \dots, \lambda; \sigma \text{ not summed}). \end{aligned}$$

Before proceeding with the problem of the second variation for the family (8.4) it will be convenient to present several definitions and prove a lemma.

Let $\eta^i(t)$ be a broken solution of the Jacobi equations which is defined and continuous for t on g , and which has corners only at the points $t = c$

and $t = a_\sigma$ ($\sigma = 1, \dots, \mu - 1, \mu + 1, \dots, \lambda$). The broken solution $\eta^i(t)$ will be termed *admissible* if it satisfies the end conditions

$$\eta^i(t') = 0, \quad \eta^i(t'') = 0, \quad (i = 1, \dots, m)$$

and if with a set (v) it satisfies the conditions

$$(8.5)' \quad \eta^i(c) = z_h^i(0)\alpha^h, \quad (h = 1, \dots, n = m - 1)$$

$$(8.5)'' \quad \eta^i(a_\sigma) = Z_h^{i\sigma}(0)\beta_\sigma^h, \quad (\sigma \text{ not summed})$$

at the corners $t = c$ and $t = a_\sigma$ respectively. If the broken solution $\eta^i(t)$ is tangential, a necessary and sufficient condition that it be admissible is that it vanish at the points at which t takes on the values

$$(8.6) \quad a_0, a_1, \dots, a_{\mu-1}, a_\mu, a_{\mu+1}, \dots, a_{\lambda+1} \quad (a_\mu = c).$$

Any two admissible broken solutions of the Jacobi equations will be said to be *equal mod T^0* , if their difference is an admissible broken tangential solution. Understanding that a solution *determined mod T^0* means a solution which is unique except for the possible addition of an admissible broken tangential solution, Lemma 8.1 is as follows.

LEMMA 8.1. *An admissible broken solution of the Jacobi equations determines a unique set (v) , and is determined mod T^0 by a set (v) .*

The first part of the lemma follows from the fact that each of the intermediate manifolds M_q is regular. To prove the second part we note that the variation $\eta^i(t) = x_e^i(t, 0)$ of the family (8.4) is an admissible broken solution corresponding to a given set (v) . Let $\bar{\eta}^i(t)$ be any other admissible broken solution corresponding to the same set (v) . Then the difference $w^i(t) = \eta^i(t) - \bar{\eta}^i(t)$ is a broken solution of the Jacobi equations which vanishes at the points at which t takes on the values (8.6). Since no one of the $\lambda + 1$ segments

$$(8.7) \quad a_j \leq t \leq a_{j+1} \quad (j = 0, 1, \dots, \lambda)$$

of g contains a conjugate point of its initial end point, $w^i(t)$ is an admissible broken tangential solution. Hence $\eta^i(t)$ and $\bar{\eta}^i(t)$ are equal mod T^0 , and the proof of the lemma is complete.

Returning to the function $J(ev)$, differentiating twice with respect to e , integrating by parts, and setting $e = 0$, we find that

$$(8.8) \quad J^0_{v^r v^i v^r v^s} = x_{ee}^i F^{1r^i} \Big|_{t'}^c + \sum_{\sigma} x_{ee}^i F^{\kappa r^i} \Big|_{a_{\sigma}^+}^{a_{\sigma}^-} + x_{ee}^i F^{2r^i} \Big|_c^{t''} \\ + \int_{t'}^c 2\Omega^1(\eta, \dot{\eta}) dt + \int_{a_{\sigma}^+}^{a_{\sigma}^-} 2\Omega^2(\eta, \dot{\eta}) dt \\ (i = 1, \dots, m; \sigma = 1, \dots, \mu - 1, \mu + 1, \dots, \lambda)$$

where $\eta^i(t)$ is the variation $x_e^i(t, 0)$ of the family (8.4) and where κ is 1 for a_σ on g_1 and 2 for a_σ on g_2 . The terms corresponding to the limits t' and t'' vanish in (8.8) as do the terms corresponding to the points a_σ . Hence we have the following theorem, readily proved with the aid of Lemma 8.1.

THEOREM 8.2. *The index form $Q(v)$ admits the representation*

$$Q(v) = b_{hk} \alpha^h \alpha^k + \int_{t'}^c 2\Omega^1(\eta, \dot{\eta}) dt + \int_c^{t''} 2\Omega^2(\eta, \dot{\eta}) dt$$

$$(h, k = 1, \dots, n = m - 1)$$

where $\eta^i(t)$ is any admissible broken solution of the Jacobi equations determined mod T^0 by (v) , and where

$$b_{hk} = [F^1_{r^i}(\gamma^-, \dot{\gamma}^-) - F^2_{r^i}(\gamma^+, \dot{\gamma}^+)] z^i_{hk}(0) \quad (i = 1, \dots, m).$$

An admissible broken solution of the Jacobi equations $\eta^i(t)$ will be termed a *special solution* if at $t = c$ it satisfies the secondary incidence relations with $\omega^h = \alpha^h$ therein and if corresponding to each corner $t = a_\sigma$ there is a constant k such that

$$\Delta \dot{\eta}^i = \dot{\eta}^i]_{a_\sigma^+}^{a_\sigma^-} = k \dot{\gamma}^i(a_\sigma) \quad (i = 1, \dots, m).$$

We shall prove the following lemma.

LEMMA 8.2. *A necessary and sufficient condition that a set $(v) \neq (0)$ be a critical point of $Q(v)$ is that (v) determine mod T^0 a special solution of the Jacobi equations.*

A necessary and sufficient condition that $(v) \neq (0)$ be a critical point of $Q(v)$ is that

$$(8.9) \quad Q_{v^r} = 0 \quad (r = 1, \dots, \lambda n).$$

We shall prove first that the conditions (8.9) imply that any admissible broken solution (η) determined mod T^0 by (v) is a special solution. For σ fixed, the n members of (8.9) representing the partial derivatives of $Q(v)$ with respect to $(\beta_\sigma^1, \dots, \beta_\sigma^n)$ may be written in the form

$$(8.10)' \quad \Delta \xi_i Z_k^{i\sigma} = 0, \quad (i = 1, \dots, m; k = 1, \dots, n)$$

where

$$\Delta \xi_i = \Omega^{\kappa}_{\eta^i}]_{a_\sigma^+}^{a_\sigma^-}$$

with $\kappa = 1$ or 2 , according as a_σ is a point on g_1 or g_2 respectively. Moreover

$$(8.10)'' \quad \Delta \xi_i \dot{\gamma}^i(a_\sigma) = 0.$$

The m equations (8.10) have a unique solution $\Delta\zeta_i = 0$. But from the definition of $\Delta\zeta_i$, we see that $\Delta\zeta_i = 0$ implies that

$$(8.11) \quad \Delta\dot{\eta}^i = k\dot{\gamma}^i(a_\sigma)$$

for σ fixed.

The n members of (8.9) representing the partial derivatives of $Q(v)$ with respect to $(\alpha^1, \dots, \alpha^n)$ may be written in the form

$$(8.12) \quad b_{hk}\alpha^h + z_k^i(0)\zeta_i]_+^- = 0 \quad (i = 1, \dots, m; h, k = 1, \dots, n).$$

Thus (η) satisfies the secondary incidence relations at $t = c$ with $\omega^h = \alpha^h$ therein, and the condition of the lemma is necessary for a critical point.

Conversely if (η) is a special solution determined mod T^0 by (v) , (8.12) holds for $t = c$, and (8.11) holds for each σ . It follows that (8.10) holds for each σ . Hence (8.9) holds, and the condition of the lemma is sufficient for a critical point.

Understanding that solutions which are *linearly independent mod T^0* are solutions which are linearly independent of admissible broken tangential solutions, the following lemma is readily proved.

LEMMA 8.3. *A set of admissible broken solutions of the Jacobi equations are linearly independent mod T^0 or not, according as the sets (v) are linearly independent or not, and conversely.*

From Lemmas 8.2 and 8.3 we infer that the nullity of the index form is equal to the number of special solutions which are linearly independent mod T^0 .

A solution of the Jacobi equations for t on g which is of class C^1 for t on g_1 and g_2 respectively and of class C^2 on each interval (8.7), which vanishes at $t = t'$ and $t = t''$, and which satisfies the secondary incidence relations at $t = c$ will be termed a *reflected solution*. A reflected solution is *admissible* if it satisfies conditions of the form (8.5)'' at each point $t = a_\sigma$. An admissible reflected solution is a special solution which has no corners at the points $t = a_\sigma$. We shall prove the following lemma.

LEMMA 8.4. *Any special solution of the Jacobi equations is identically equal mod T^0 to an admissible reflected solution.*

Let $\eta^i(t)$ be any special solution. Let $\rho(t)$ be a continuous function for t on g with the following properties. On each interval (8.7), $\rho(t)$ is of class C^2 with $\rho(a_j) = \rho(a_{j+1}) = 0$, and with $\dot{\rho}(a_j)$ and $\dot{\rho}(a_{j+1})$ so chosen that for t on g , the solution

$$\bar{\eta}^i(t) \equiv \eta^i(t) - \rho(t)\dot{\gamma}^i(t) \quad (i = 1, \dots, m)$$

has no corners at the points $t = a_\sigma$ ($\sigma = 1, \dots, \mu - 1, \mu + 1, \dots, \lambda$). Then $\bar{\eta}^i(t)$ is an admissible reflected solution, and the proof of the lemma is complete.

With (η) and $(\bar{\eta})$ defined as in the proof of Lemma 8.4, we see that $(\bar{\eta})$ is identically equal to an admissible tangential reflected solution if and only if $(\eta) \equiv (0), \text{ mod } T^0$. Moreover, if a finite set S of special solutions be replaced by a set A of admissible reflected solutions, equal mod T^0 respectively to the special solutions of S , then the members of A are linearly independent of admissible tangential reflected solutions if and only if the members of S are linearly independent mod T^0 .

We shall prove the following theorem.

THEOREM 8.3. *The nullity of the index form $Q(v)$ equals the order of t'' as a conjugate point of t' .*

If the nullity of $Q(v)$ is ν , there are ν special solutions which are linearly independent mod T^0 , and therefore ν admissible reflected solutions which are linearly independent of admissible tangential reflected solutions. It follows from Theorem 7.6 that t'' is a conjugate point of t' of order ν .

On the other hand, if t'' is a conjugate point of t' of order ν , there are ν reflected solutions

$$\eta_k^i(t) \quad (i = 1, \dots, m; k = 1, \dots, \nu)$$

which are linearly independent of tangential reflected solutions. The solutions $\eta_k^i(t)$ are not necessarily admissible in that they may not satisfy conditions of the form (8.5)'' at the points $t = a_\sigma$ ($\sigma = 1, \dots, \mu - 1, \mu + 1, \dots, \lambda$).

But any reflected solution can be made admissible by adding a suitably chosen tangential reflected solution. Moreover, if a finite set R of reflected solutions be replaced by a set A of admissible reflected solutions which are equal, modulo a tangential reflected solution, respectively to the members of R , then the members of A are linearly independent of admissible tangential reflected solutions if and only if the members of R are linearly independent of tangential reflected solutions.

We assume then, that the ν reflected solutions $\eta_k^i(t)$ which are linearly independent of tangential reflected solutions have been replaced by ν admissible reflected solutions $\bar{\eta}_k^i(t)$ which are equal, modulo a tangential reflected solution, respectively to the solutions $\eta_k^i(t)$. It follows that the solutions $\bar{\eta}_k^i(t)$ are linearly independent of admissible tangential reflected solutions. But the ν admissible reflected solutions $\bar{\eta}_k^i(t)$ are special solutions which are of class C^1 for t on g_1 and g_2 respectively. Hence the nullity of $Q(v)$ is ν .

We continue with the following theorem.

THEOREM 8.4. *The index of $Q(v)$ equals the sum of the orders of the conjugate points of t' on $t' < t < t''$.*

To prove Theorem 8.4 we use the method of Morse (2, Th. 4.2). That is, we replace the extremaloid g by the subarc g_b on which $t' \leq t \leq b$ where $t' < b \leq t''$. If $t' < b \leq c$, then g_b is an extremal whereas if $c < b \leq t''$, g_b is an extremaloid. On g_b we introduce λ intermediate manifolds as previously with

$$(8.13) \quad a_0 < a_1 < \cdots < a_{\lambda+1} = b,$$

and with the points $t = a_q$ ($q = 1, \cdots, \lambda$) so distributed on g_b that no one of the $\lambda + 1$ segments into which g_b is thereby divided contains a conjugate point of its initial end point. For $b > c$, the point $t = c$ must be taken as one of the admissible points $t = a_q$, and the deflecting manifold used as the corresponding intermediate manifold.

The family of broken extremals which hereby replaces the family $X^i(t, v)$ is denoted by $X^i(t, v, b)$. The functions $J^b(v)$ and $Q^b(v)$ are defined for g_b as were $J(v)$ and $Q(v)$ for g .

The proof of Theorem 8.4 will be based on the following statement:

(α). *For any point b the index of $Q^b(v)$ is equal to the sum of the orders of the conjugate points of t' on $t' < t < b$.*

Morse proved that statement (α) is true for b on g_1 ; that is, for $t' < b \leq c$. It remains to prove that (α) is true for $c < b \leq t''$.

As b increases, the index of $Q^b(v)$ [written $I(Q^b)$] will change at most when b passes through a conjugate point a of t' and will then increase by at most the order ν of a as a conjugate point of t' . Hence for each value of $b > c$, we have

$$(8.14) \quad I(Q^b) \leq \sum \nu_a \quad (t' < a < b).$$

We shall prove the following lemma.

LEMMA 8.5. *For $b > c$ and nearer to c than any conjugate point of t' , excepting possibly c itself,*

$$(8.15) \quad I(Q^b) \geq \sum \nu_a \quad (t' < a < b).$$

With any admissible set (8.13) for which $b = a_{\lambda+1}$ satisfies the hypothesis of the lemma and $a_\lambda = c$, we proceed as follows: We denote $a_{\lambda+1}$ by $a_{\lambda+2}$ or b , and a_λ by $a_{\lambda+1}$ or c , and insert a new point a_λ between $a_{\lambda-1}$ and c . We introduce a new intermediate manifold M_λ cutting g_b at a_λ but not tangent to g_b at a_λ . For this construction we replace the set of parameters (v) by a set of $(\lambda + 1)n$ parameters (ξ) , the first $(\lambda - 1)n$ and the last n of which form

the set (v) . The broken extremal $\mathcal{E}(\xi)$ determined by (ξ) and the end points of g_b coincides with the broken extremal $X^i(t, v, b)$ for $t' \leq t \leq a_{\lambda-1}$ and $c \leq t \leq b$.

Understanding that $J_{\cdot b}(\xi)$ and $Q_{\cdot b}(\xi)$ denote the functions replacing $J^b(v)$ and $Q^b(v)$, we shall first prove that

$$(8.16) \quad I(Q^b) \geq I(Q_{\cdot b}).$$

To that end let e be a parameter near 0 and set

$$\phi(e) = J_{\cdot b}(e\xi) - J^b(ev).$$

When the first $(\lambda - 1)n$ and the last n components of (ξ) are given by (v) , the inequality $\phi(e) \geq 0$ holds for e sufficiently near 0, and $\phi(e)$ has a minimum for $e = 0$. Hence $\phi''(0) \geq 0$, and (8.16) follows with the aid of Lemma 3.1a of Morse 2.

Next we set the last n variables of (ξ) equal to zero in $Q_{\cdot b}(\xi)$, and obtain thereby a quadratic form Q_0^c . Applying Lemma 3.2b of Morse 2 we see that

$$(8.17) \quad I(Q_{\cdot b}) \geq I(Q_0^c) + N(Q_0^c),$$

where $I(Q_0^c)$ and $N(Q_0^c)$ denote the index and nullity respectively of Q_0^c . But since the nullity of Q_0^c is equal to the order of c as a conjugate point of t' and since the index of Q_0^c is equal to the sum of the orders of the conjugate points of t' on $t' < t < c$, the inequality (8.15) follows from (8.16) and (8.17), and the proof of the lemma is complete.

Upon comparing the inequalities (8.14) and (8.15), we see that for $b > c$, and sufficiently near c , statement (α) is true.

That statement (α) holds for any point $b > c$ on g_2 can be proved by taking an arbitrary point t^0 , where $c < t^0 \leq t''$, and showing (1) that if (α) is true for $b < t^0$, then it is true for $b = t^0$ and (2) that if (α) is true for $b \leq t^0$, then it is true for $b > t^0$. The method of proof is similar, and the details will be omitted. See Morse 2, Lemmas C and D.

The index and nullity of $Q(v)$ are independent of the number, position and representation of the intermediate manifolds, provided they are admissibly distributed and represented, and one intermediate manifold coincides with the deflecting manifold. The index and nullity of $Q(v)$ depend only on the conjugate points of t' on g , and their orders.

Recall that the nullity of a critical point of a function of a finite number of variables is defined as the nullity of the Hessian of the function at the critical point, and that the index of a critical point is defined as the number of negative characteristic roots of the Hessian of the function at the critical point. Understanding that each conjugate point is to be counted a

number of times equal to its order, we summarize our results in the Index Theorem.

INDEX THEOREM. *The point $(v) = (0)$ is a critical point of $J(v)$ with an index equal to the number of conjugate points of $t = t'$ on g preceding $t = t''$, and a nullity equal to the order of $t = t''$ as a conjugate point of $t = t'$.*

Let $t = a$ and $t = b$ be any two points on g . We shall prove the following theorem.

THEOREM 8.5. *The numbers of zeros of the two conjugate point determinants $D_g(t, a)$ and $D_g(t, b)$ on any finite open interval of g differ by at most n , where $n = m - 1$.*

Let $p < t < q$ be any finite open interval of g . Let r be a point following p on $p < t < q$. Suppose r is not a or b or c and that there is no conjugate point of a or b on $p < t \leq r$. Similarly let s be a point preceding q on $r < t < q$. Suppose s is not a or b or c and that there is no conjugate point of a or b on $s \leq t < q$. Understanding that $Q(rs)$ denotes the index form corresponding to the finite open interval $r < t < s$ of g , and that $I(rs)$ denotes the index of $Q(rs)$, we shall first establish the following statement.

(α). *The number of zeros of $D_g(t, a)$ on $r < t < s$ is equal to*

$$(8.18) \qquad I(rs) + k \qquad (0 \leq k \leq n)$$

where k is an integer or zero.

There are three cases to be considered:

Case 1. $a < r < s$,

Case 2. $r < s < a$,

Case 3. $r < a < s$.

To prove Case 1 we first set up index forms $Q(ar)$ and $Q(rs)$. We then set up $Q(as)$ taking one intermediate manifold, M_r , at the point r and the same intermediate manifolds preceding and following M_r as are used to define $Q(ar)$ and $Q(rs)$ respectively. When the variables of $Q(as)$ belonging to the intermediate manifolds preceding and following M_r are the same as the variables of $Q(ar)$ and $Q(rs)$ respectively, the quadratic form obtained by setting the n variables belonging to M_r in $Q(as)$ equal to zero is equal to $Q(ar) + Q(rs)$. Hence

$$I(as) - n \leq I(ar) + I(rs) \leq I(as).$$

See Morse 1, p. 62, Lemma 7.2. It follows that

$$(8.19) \quad I(as) - I(ar) = I(rs) + k, \quad (0 \leq k \leq n)$$

where k is an integer or zero. But, since conjugate points are counted according to their orders, the left member of (8.19) represents the number of conjugate points of a on $r < t < s$. Moreover, since the order of a conjugate point of a is equal to the order of vanishing of $D_g(t, a)$ at that point, the number of zeros of $D_g(t, a)$ on $r < t < s$ is given by (8.18) when $a < r < s$.

For Case 2 we interchange the rôles of r and s in Case 1 and obtain

$$I(ar) - I(as) = I(sr) + k, \quad (0 \leq k \leq n)$$

where k is an integer or zero. But $I(sr) = I(rs)$, and statement (α) follows as in Case 1.

For Case 3 we first set up index forms $Q(ra)$ and $Q(as)$, and then $Q(rs)$, taking one intermediate manifold, M_a , at a to define $Q(rs)$, and taking the same intermediate manifolds preceding and following M_a as are used for $Q(ra)$ and $Q(as)$ respectively. If in particular $a = c$, then M_a must be taken as the deflecting manifold M . As in Case 1 we have

$$I(rs) - n \leq I(ra) + I(as) \leq I(rs).$$

Since $I(ar) = I(ra)$, it follows that

$$(8.20) \quad I(ar) + I(as) = I(rs) - n + k, \quad (0 \leq k \leq n)$$

where k is an integer or zero. The number of conjugate points of a on $r < t < s$ is given by the left member of (8.20). But since the conjugate point determinant $D_g(t, a)$ vanishes to the n -th order at a , and since $r < a < s$, the number of zeros of $D_g(t, a)$ on $r < t < s$ is given by the right member of (8.20) increased by n . This completes the proof of statement (α).

Returning to the theorem, we see that the numbers of zeros of $D_g(t, a)$ and $D_g(t, b)$ on $r < t < s$ differ by at most n . From our choice of r and s , it follows that the numbers of zeros of $D_g(t, a)$ and $D_g(t, b)$ on $p < t < q$ differ by at most n .

The following is an easy corollary.

COROLLARY 8.5. *The numbers of zeros of two conjugate point determinants $D_g(t, a)$ and $D_g(t, b)$ on any finite interval (open or closed) of g differ by at most n , where $n = m - 1$.*

Conclusion. The index theory can be directly extended to the case that g is a broken extremal with any finite number of corners, at each of which g

is cut across by a regular $(m-1)$ -manifold of class C^2 , not tangent to either arc of g at the corner, and at each of which g satisfies a corresponding set of primary incidence relations. Moreover, if the initial end point t' of a broken extremal g lies on a regular $(m-1)$ -manifold \mathcal{M} of class C^2 , which cuts g transversally at t' , but is not tangent to g at t' , the index theory has corresponding theorems, stated in terms of "focal" points of \mathcal{M} and their orders.

SWEET BRIAR COLLEGE,
SWEET BRIAR, VIRGINIA.

BIBLIOGRAPHY.

Bolza, O.

1. *Vorlesungen über Variationsrechnung*, Berlin, Teubner (1909).

Bliss, G. A.

1. (With Mason, M.) "A problem of the calculus of variations in which the integrand is discontinuous," *Transactions of the American Mathematical Society*, vol. 7 (1906), pp. 325-336.
2. (With Mason, M.) "The properties of curves in space which minimize a definite integral," *Transactions of the American Mathematical Society*, vol. 9 (1908), pp. 440-466.
3. "Jacobi's condition for problems of the calculus of variations in parametric form," *Transactions of the American Mathematical Society*, vol. 17 (1916), pp. 195-206.

Graves, L. M.

1. "Discontinuous solutions in space problems of the calculus of variations," *American Journal of Mathematics*, vol. 52 (1930), pp. 1-28.
2. "Discontinuous solutions in the calculus of variations," *Bulletin of the American Mathematical Society*, vol. 36 (1930), pp. 831-846. This paper contains further references.

Miles, E. J.

1. "Some properties of space curves minimizing a definite integral with discontinuous integrand," *Bulletin of the American Mathematical Society*, vol. 20 (1913), pp. 11-19.

Morse, M.

1. "The calculus of variations in the large," *American Mathematical Society Colloquium Publications* 18, New York (1934).
2. "The index theorem in the calculus of variations," *Duke Mathematical Journal*, vol. 4 (1938), pp. 231-246.

ON 0-REGULAR TRANSFORMATIONS.*

By A. D. WALLACE.

1. Introduction. In this paper we consider a particular type of interior transformation which we call a 0-regular transformation. A mapping of this type may be roughly described by saying that the inverse sets (of points) are uniformly locally connected and in addition form a continuous collection. More accurately we require of the continuous transformation $T(A) = B$ that for any sequence $b_n \rightarrow b$ in B we have (i) $T^{-1}(b_n) \rightarrow T^{-1}(b)$ and (ii) this convergence be regular relative to 0-cycles in the sense of Whyburn. The condition (i) is a characterization of interior transformations due to Eilenberg. There are many obvious generalizations of this notion with which we shall not be concerned.

We show that any 0-regular transformation may be factored into two 0-regular transformations of which the first is monotone and the second of constant multiplicity. This result is important in studying the effect of the transformation. It is also shown that 0-regular convergence is preserved under the inverse of a 0-regular transformation. We prove that (the mapped space being a locally connected continuum) cut-points, end-points and A -sets map respectively into cut-points, end-points and A -sets. In particular a 0-regular transformation is topological on a dendrite, and is monotone if the image space is a dendrite.

2. General theorems. We suppose throughout that $T(A) = B$ is a continuous transformation defined on the metric space A , and that B contains more than one point. The following definition is due to G. T. Whyburn [1]: If the sequence of closed sets $\{M_n\}$ converges to M , then $M_n \rightarrow M$ 0-regularly provided that for each $\epsilon > 0$ there are positive numbers δ and N such that for $n > N$ any two points x and y in M_n with $\rho(x, y) < \delta$, lie in an ϵ -continuum¹ in M_n . It is readily seen that the following result holds:

(2.1) *If $M_n \rightarrow M$ in a compact metric space, then in order that the convergence be 0-regular it is necessary and sufficient that for each positive ϵ there exist positive numbers δ and N such that for $p \in M$ and $n > N$, the set $V_\delta(p) \cdot M_n$ is contained in a connected subset of $V_\epsilon(p) \cdot M_n$.*²

* Received July 31, 1939.

¹ An ϵ -set is a set of diameter less than ϵ .

² The symbol $V_\epsilon(p)$ denotes the set of points not farther from p than ϵ .

The following example is of interest: Let $\{f_n(x)\}$ be a sequence of real-valued continuous functions defined on the unit interval and converging to the function $f_0(x)$. Let M_i be the graph of the function $y = f_i(x)$. In order that $f_n(x) \rightarrow f_0(x)$ uniformly it is necessary and sufficient that $M_n \rightarrow M_0$ 0-regularly.

I shall say that the transformation $T(A) = B$ is 0-regular provided that if $y_n \rightarrow y$ in B , the sets $T^{-1}(y_n) \rightarrow T^{-1}(y)$ 0-regularly. It follows immediately from a theorem due to Eilenberg [2] that if T is 0-regular it is interior; that is, open sets map into open sets [3]. The proof of the following result presents no difficulty:

(2.2) *In order that the interior transformation $T(A) = B$ be 0-regular, where A is compact, it is necessary and sufficient that for each $\epsilon > 0$ there exist a $\delta > 0$, such that if x and y are in A with $\rho(x, y) < \delta$ and $T(x) = T(y)$, then x and y lie in an ϵ -continuum in $T^{-1}T(x) = T^{-1}T(y)$.*

(2.3) *If $T(A) = B$ is interior, where A is compact, and if the sequence of closed sets $Y_n \rightarrow Y$ in B , then we have $T^{-1}(Y_n) \rightarrow T^{-1}(Y)$, and if this convergence is 0-regular so is the convergence $Y_n \rightarrow Y$.*

Proof. For the proof that $T^{-1}(Y_n) \rightarrow T^{-1}(Y)$ see [4]. Assume that the convergence is 0-regular. Let ϵ be a positive number and e a positive number such that if $a \in A$ and $b = T(a)$ then $T(V_e(a)) \subset V_\epsilon(b)$; take $d > 0$ and $N > 0$ for this e as in (2.1). Pick $\delta > 0$ so that if $b \in B$ and $a \in T^{-1}(b)$ we have $V_\delta(b) \subset T(V_d(a))$. This latter is possible by a theorem due to G. T. Whyburn [5]. Let p be any point of Y and let x and y be points of $V_\delta(p) \cdot Y_n$, $n > N$. If $q \in T^{-1}(p) \subset T^{-1}(Y)$, then $V_\delta(p) \cdot Y_n \subset T(V_d(q) \cdot T^{-1}(Y_n))$ and we can find points x' and y' in $V_d(q) \cdot T^{-1}(Y_n)$ mapping into x and y respectively. Since $n > N$ we know that x and y lie in a connected subset H of $V_e(q) \cdot T^{-1}(Y_n)$ in virtue of the 0-regular convergence $T^{-1}(Y_n) \rightarrow T^{-1}(Y)$. Hence $T(H) \subset T(V_e(q) \cdot T^{-1}(Y_n)) \subset V_\epsilon(p) \cdot Y_n$. Thus x' and y' lie in a connected subset of $V_\epsilon(p) \cdot Y_n$. This completes the proof in virtue of (2.1)

(2.31) *If $T(A) = B$ is 0-regular and A is compact, and T is factored, $T = T_2 T_1$, so that $T_1(A) = A'$ is interior, then $T_2(A') = B$ is 0-regular.*

Proof. Suppose that $y_n \rightarrow y$ in B . Then $T^{-1}(y_n) \rightarrow T^{-1}(y)$ 0-regularly in A . Hence $T_1 T^{-1}(y_n) \rightarrow T_1 T^{-1}(y)$ 0-regularly by (2.3); but for any $b \in B$ we have $T_1 T^{-1}(b) = T_2^{-1}(b)$, so that $T_2^{-1}(y_n) \rightarrow T_2^{-1}(y)$ 0-regularly.

The following example shows that this theorem is false if T_1 is not interior: Let A be the circle $|z| = 1$ and B the circle $|w| = 1$ and T the

transformation $w = z^2$. Let T_1 be the transformation mapping A into a lemniscate by identifying the points 1 and -1 . Then T_2 is not 0-regular.

Before proceeding to the proof of a factor theorem we need the following

LEMMA. *Let $\{M_n\}$ be a sequence of disjoint locally connected closed sets converging 0-regularly to the locally connected set M in a compact metric space. Suppose that $M \cdot M_n = 0$. If $M = X^1 + \cdots + X^k$ is a decomposition into components then there is an integer N such that if $n > N$, $M_n = X_n^1 + \cdots + X_n^k$ is a decomposition into components and for each $i = 1, 2, \dots, k$, we have $X_n^i \rightarrow X^i$ 0-regularly and if $i \neq j$, $X^i \cdot \lim X_n^j$ is vacuous.*

Proof. Let X be a component of M and $\{X_n\}$ a sequence of components of the sets $\{M_n\}$ so chosen that X and $\liminf X_n$ have a point in common. Let $\{X_{n_i}\}$ be a convergent subsequence, say $X_{n_i} \rightarrow X'$. From the local connectivity of M_n it follows that $M_n - X_n$ is closed. Let $\{M_{n_i} - X_{n_i}\}$ be a convergent sequence chosen from the sequence $\{M_{n_i} - X_{n_i}\}$ and converging to a set Y . It follows that $M = Y + X'$ and since $(M_{n_i} - X_{n_i}) \cdot X_{n_i} = 0$, we have $Y \cdot X' = 0$ as a consequence of the 0-regular convergence [1]. Since $X \cdot X' \neq 0$ and X' is a continuum it follows that $X = X'$. Thus every convergent subsequence of $\{X_n\}$ converges to X , so that $X = \lim X_n$.

It follows immediately from the definition of 0-regular convergence that $X \limsup (M_n - X_n)$ is empty so that $\limsup (M_n - X_n) \subset M - X$; from the fact that $X_n \rightarrow X$, a component of M , we deduce that $M - X \subset \liminf (M_n - X_n)$. Hence $M_n - X_n \rightarrow M - X$. Since $(M_n - X_n) \cdot X_n = 0$ and $(M - X) \cdot X = 0$ we conclude that the former sets converge 0-regularly to the latter and thus [1] that $X_n \rightarrow X$ and $M_n - X_n \rightarrow M - X$ 0-regularly. The proof of the lemma may be carried through by induction.

The transformation $T(A) = B$ is *locally topological* or a *local homeomorphism*, provided that it is interior and that each point in A admits a neighborhood on T is topological [6]; T is said to be *monotone* provided that for each $b \in B$, the set $T^{-1}(b)$ is connected [7].

(2.4) *If $T(A) = B$ is 0-regular and A is a continuum, then T can be factored $T = T_2 T_1$, so that*

- (i) $T_1(A) = A'$ is monotone and 0-regular
- (ii) $T_2(A') = B$ is of constant multiplicity and locally topological.

Proof. As a consequence of a theorem due to Whyburn [8] we know that T can be factored so that T_1 is monotone and T_2 is interior. Also for each $b \in B$ it follows that $T^{-1}(b)$ is locally connected. If $y_n \rightarrow y$ in A' we must

show that $T_1^{-1}(y_n) \rightarrow T_1^{-1}(y)$ 0-regularly. But from the proof of the factor theorem just cited it follows that each set $T_1^{-1}(x)$, $x \in A'$, is a component of $T^{-1}T_2(x)$ and since T is 0-regular $T^{-1}T_2(y_n) \rightarrow T^{-1}T_2(y)$. But T_1 is continuous so that $T_1^{-1}(y) \limsup T_1^{-1}(y) \neq 0$, so that by the lemma $T_1^{-1}(y_n) \rightarrow T_1^{-1}(y)$ 0-regularly. This completes the proof of (i).

By (2.31) T_2 is 0-regular since we have shown that T_1 is 0-regular and hence interior. Let $p(b)$ be the number of points in $T_2^{-1}(b)$, that is, the number of components in $T^{-1}(b)$. By the lemma $p(b)$ is a continuous function and since B is connected it follows that $p(b)$ is constant. It readily follows from this and the 0-regularity of T_2 that this transformation is locally topological.

As a matter of convenience we state explicitly the following:

(2.41) *If $T(A) = B$ is 0-regular on the continuum A and X is a component of $T^{-1}(y)$ then there exists a sequence of points $y_n \rightarrow y$ in B and a sequence of components $\{X_n\}$ of $\{T^{-1}(y_n)\}$ converging 0-regularly to X . The sequence $\{X_n\}$ is essentially unique in the sense that if $\{Z_n\}$ is any sequence of components of $\{T^{-1}(y_n)\}$ converging to X , then the sequences $\{X_n\}$ and $\{Z_n\}$ differ only in a finite number of terms.*

(2.5) *In order that the transformation $T(A) = B$ be 0-regular on the compact space A , it is necessary and sufficient that for any sequence of closed sets $Y_n \rightarrow Y$ 0-regularly we have $T^{-1}(Y_n) \rightarrow T^{-1}(Y)$ 0-regularly.*

Proof. The sufficiency of the condition is obvious. Assume now that T is 0-regular and let $Y_n \rightarrow Y$ 0-regularly. Since T is interior it follows that $T^{-1}(Y_n) \rightarrow T^{-1}(y)$ [4]. Let $\epsilon > 0$ and select $u > 0$ from (2.2) so that if z and w are any two points of A such that $T(z) = T(w)$ then z and w lie in an $\epsilon/3$ continuum in $T^{-1}T(z)$. Let t be a positive number less than the smaller of $u/3$ and $\epsilon/3$. Let $e > 0$ be so chosen that if $b \in B$ and $a \in T^{-1}(b)$, then $V_e(b) \subset T(V_t(a))$ [8]. Since $Y_n \rightarrow Y$ 0-regularly there are positive numbers d and N such that if $n > N$ and $q \in Y$, then any two points of $V_d(q) \cdot Y_n$ lie in a connected subset of $V_e(q) \cdot Y_n$, by (2.1). Since T is continuous and A is compact there is a positive number δ such that if $a \in A$ and $b = T(a)$ then $T(V_\delta(a)) \subset V_d(b)$. If $n > N$ and $p \in T^{-1}(Y)$ I have to show that any two points x and y in $V_\delta(p) \cdot T^{-1}(Y_n)$ lie in a continuum in $V_e(p) \cdot T^{-1}(Y_n)$. To this end it is shown that x and y can be r -chained for all small positive r in the set $V_{2\epsilon/3}(p) \cdot T^{-1}(Y_n)$.

Let r be positive and less than $u/3$ and pick $s > 0$ so that if $b \in B$ and $a \in T^{-1}(b)$ then $V_s(b) \subset T(V_r(a))$. Since x and y are in $V_\delta(p) \cdot T^{-1}(Y_n)$ then x' and y' (the images of x and y respectively) are in $V_d(p')$, $p' = T(p)$.

It follows that $x' + y' \subset$ a connected subset of $V_e(p') \cdot Y_n$. Hence we can find a chain

$$x' = b_0, b_1, \dots, b_k = y', \quad b_j \in V_e(p') \cdot Y_n, \quad \rho(b_j, b_{j+1}) < s.$$

Now $V_e(p') \cdot Y_n \subset T(V_t(p) \cdot T^{-1}(Y_n))$ so that we can find points

$$x = a_0, a_1, \dots, a_k = y, \quad a_j \in V_t(p) \cdot T^{-1}(Y_n), \quad T(a_j) = b_j.$$

Since

$$b_{j+1} \in V_s(b_j) \cdot Y_n \subset T(V_r(a_j) \cdot T^{-1}(Y_n))$$

there is a point c_{j+1} in $V_r(a_j) \cdot T^{-1}(Y_n)$ mapping onto b_{j+1} . For each j we thus have $\rho(a_j, c_{j+1}) < r$. Also

$$\rho(a_{j+1}, c_{j+1}) \leq \rho(a_j, c_{j+1}) + \rho(a_j, p) + \rho(a_{j+1}, p) < r + 2t < u.$$

Hence a_{j+1} and c_{j+1} lie in a continuum K_{j+1} of diameter less than $\epsilon/3$ in $T^{-1}T(a_{j+1}) \subset T^{-1}(Y_n)$. Since $\rho(a_{j+1}, p) < \epsilon/3$, no point of K_{j+1} is farther from p than $2\epsilon/3$. We can now chain a_{j+1} to c_{j+1} by an r -chain in $V_{2\epsilon/3}(p) \cdot T^{-1}(Y_n)$. Thus x can be r -chained to y for all small r in $V_{2\epsilon/3}(p) \cdot T^{-1}(Y_n)$ and it follows that x and y lie in a connected subset of $V_\epsilon(p) \cdot T^{-1}(Y_n)$.³

It is clear that (2.5) implies the following:

(2.51) *If Y is a locally connected closed set in B then $T^{-1}(Y)$ is locally connected.*

From (2.5) we also get the following product theorem:

(2.6) *If $T_1(A) = A'$ and $T_2(A') = B$ are 0-regular where A is compact, then so also is $T = T_2T_1$.*

Proof. For if $y_n \rightarrow y$ in B , then $T_2^{-1}(y_n) \rightarrow T_2^{-1}(y)$ 0-regularly. Hence $T_1^{-1}T_2^{-1}(y_n) \rightarrow T_1^{-1}T_2^{-1}(y)$ 0-regularly. But for any $b \in B$ we have $T^{-1}(b) = T_2^{-1}T_1^{-1}(b)$.

3. 0-Regular transformations on Continua. In this section we shall suppose that $T(A) = B$ is 0-regular and that A is a continuum.

(3.1) *If H is any subset of B then T is 0-regular on $T^{-1}(H)$. If H is a connected subset of B then T is 0-regular on each of the finite number of components of $T^{-1}(H)$ and each such component maps onto all of H under T .*

³This proof is similar to proofs given by G. T. Whyburn [5] and W. T. Puckett [11] for somewhat different results. The result (2.51) has also been proved in the cited paper of Puckett.

Proof. The proof of the first statement is immediate. Assume that H is a connected subset of B and $T^{-1}(H) = P + Q$ where $\bar{P}Q + P\bar{Q} = 0$. It is readily seen that if $P \neq 0$ we have $T(P) = H$ and since for each $b \in B$ the set $T^{-1}(b)$ has only a finite number of components, there are at most a finite number of components in $T^{-1}(H)$. Since any one of these can be taken as P in the above decomposition it only remains to show that T is 0-regular on each component. If K is a component of $T^{-1}(H)$ then K is open in $T^{-1}(H)$ and hence T is interior on K . Thus if $y_n \rightarrow y$ in H we have $K \cdot T^{-1}(y_n) \rightarrow K \cdot T^{-1}(y)$. But if $b \in H$ then any component of $T^{-1}(b)$ which intersects K certainly lies in K . Hence each component of $K \cdot T^{-1}(b)$ is also a component of $T^{-1}(b)$. By (2.41) the result follows.

(3.2) *If the continuum X separates A and lies in the inverse of a point $x \in B$, then X is a component of $T^{-1}(x)$.*

Proof. We may write

$$A = H + K, \quad H \cdot K = X,$$

where H and K are continua. If $y \in B - x$, then clearly any component of $T^{-1}(y)$ is either in $H - X$ or $K - X$. Let X' be the component of $T^{-1}(x)$ containing X . In H we can select a sequence of points $\{z_n\}$ not in X and which converge to a point $z \in X$. If X_n is the component of $T^{-1}T(z_n)$ which contains z_n then X_n is in $H - X$ and by (2.41) $X_n \rightarrow X'$. Hence $X' \subset H$. Similarly $X' \subset K$, so that $X = X'$.

(3.3) *If the closed set X separates A irreducibly between two points and is contained in a component of the inverse of a point, then X is this component.*

(3.4) *If the point x of A is an end-point, a regular point in the sense of Menger, or a cut-point of A , then x is a component of $T^{-1}T(x)$, and T is locally topological in a neighborhood of x .*

Proof. The first two cases follow because x cannot lie on a continuum of convergence. The third follows from (3.2)

(3.5) *If the continuum A does not contain uncountably many non-generate mutually exclusive continua, then each 0-regular transformation on A is locally topological.*

Proof. If we assume that the result is false we can find a non-degenerate component X of the inverse of a point $x \in B$, and a neighborhood U of X such that for each $y \in B$, any component of $T^{-1}(y)$ which intersects U is non-degenerate. But $T(U)$ is a neighborhood of x and since B is a continuum

$T(U)$ contains uncountably many points. Hence A contains uncountably many non-degenerate mutually exclusive continua.

4. Results for locally connected continua. We now suppose that A is a locally connected continuum and T is 0-regular *unless the contrary is explicitly stated*. We also assume a knowledge of the cyclic element theory of such spaces, Kuratowski and Whyburn [9].

(4.1) *If T is a local homeomorphism and J is a simple closed curve in B , then each component of $T^{-1}(J)$ is a simple closed curve mapping onto all of J under T . If D is a dendrite in B then there k components in $T^{-1}(D)$ (k being the multiplicity of T) each of which is a dendrite mapping topologically onto D under T .*

Proof. If Z is a locally connected continuum and Z_1 is a component of $T^{-1}(Z)$ then $T(Z_1) = Z$ is locally topological and hence Z_1 is a locally connected continuum. The proof of the first statement is immediate. Let D_1 be a component of $T^{-1}(D)$. Since $T(D_1) = D$ is interior there exists a dendrite $D' \subset D_1$ such that $T(D') = D$ is topological by a theorem due to Whyburn [10]. But since T is locally topological it is clear that $D' = D_1$.

(4.11) *If A is a dendrite then T is topological, if B is a dendrite then T is monotone.*

Proof. Using the notation of (2.4), if A is a dendrite then A' is a dendrite since T_1 is monotone. As in (4.1) it follows that T_2 is topological. Hence T is monotone. But by (3.4) each point $x \in A$ is a component of $T^{-1}T(x)$. Similar reasoning applies to the second statement.

(4.2) *If x is a cut-point of A then $y = T(x)$ is a cut-point of B .*

Proof. By (3.4) and with the notation of (2.4) we see that x is a component of $T_1^{-1}T_1(x)$, i. e., $x = T_1^{-1}T_1(x)$ since T_1 is monotone. If $y_1 = T_1(x)$ did not cut A' then $A' - y_1$ would be connected and hence so would $T_1^{-1}(A' - y_1) = A - x$. Hence y_1 cuts A' . Thus we may write $A' = M + N$, $MN = y_1$, where M and N are non-degenerate continua. Now $T_2(y_1) = y_2$ is clearly not an end-point and if it is not a cut-point it lies in a true cyclic element E of B . We can find arcs ay_1 and by_1 in M and N respectively such that T_2 is topological on the arc $ay_1 + y_1b$. Let $T_2(a) = c$, $T_2(b) = d$. If $cy_2 - y_2$ were in $B - E$ it would lie in some component R of this set. But then clearly we would have $F(R) = \bar{R} - R = y_2$ and hence y_2 is a cut-point, contrary to our assumption. There are thus some subarcs of cy_2 and y_2d in E and we may assume that $cy_2 + y_2d$ is a subset of E . Now

c and d lie on a simple closed curve J' in E , $J' = cpd + cqd$. Let cpd be the arc of J' not containing y_2 . We may assume that c is the first point on y_2c in cpd and that d is the first point on y_2d in cpd . Finally let J be the simple closed curve $cy_2d + cpd$, and let J_1 and J_2 be the components of $T^{-1}(J)$ containing ay_1 and y_1b respectively. Now J_1 and J_2 are simple closed curves by (4.1) and since no simple closed curve can have points other than y_1 in both M and N it follows that J_1 and J_2 are different. Hence two different components of $T^{-1}(J)$ have the point y_1 in common. This is a contradiction.

(4.3) If H is an A -set in A then $T(H)$ is an A -set in B .

Proof. Let r_1s_1 be an arc in A' with r_1 and s_1 in $H_1 = T_1(H)$. Since $T_1^{-1}(r_1s_1)$ is a locally connected continuum we can find an arc rs in this set with r and s in H . Since H is an A -set rs lies in H and hence $T(rs) = r_1s_1$ lies in H_1 . Let $H_2 = T_2(H_1)$ and assume that H_2 is not an A -set. We can then find an arc $u_2x_2v_2$ in B with $u_2x_2v_2 \cap H_2 = u_2 + v_2$. There is an arc $u_2y_2v_2$ in H_2 . Let $J = u_2x_2v_2 + u_2y_2v_2$ and let y_1' be a point of $H_1 \cap T_2^{-1}(y_2)$. We can find a neighborhood U of y_1 on which T_2 is topological and $V = T_2(U)$ will be a neighborhood of y_2 . Now y_2 is in V and hence a subarc t_2 of $u_2y_2v_2$ containing y_2 lies in V . We can thus find an arc t_1 of H_1 which maps topologically onto t_2 . Let J_1 be the component of $T^{-1}(J)$ which contains t_1 . Since H_1 is an A -set it is clear that J_1 is in H_1 and hence we have J in H . Thus $u_2x_2v_2$ is in H .

THE UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY

1. G. T. Whyburn, *Fundamenta Mathematicae*, vol. 35 (1935), p. 408.
2. S. Eilenberg, *Fundamenta Mathematicae*, vol. 22 (1934), p. 292.
3. Stoilow, *Principes topologiques de la théorie des fonctions*, Paris, 1938.
4. A. D. Wallace, *American Journal of Mathematics*, vol. 61 (1939), p. 757.
5. G. T. Whyburn, *Duke Mathematical Journal*, vol. 4 (1938), p. 1.
6. S. Eilenberg, *Fundamenta Mathematicae*, vol. 24 (1935), p. 160.
7. G. T. Whyburn, *American Journal of Mathematics*, vol. 56 (1934), p. 370.
8. G. T. Whyburn, *Duke Mathematical Journal*, vol. 3 (1937), p. 370.
9. Kuratowski and Whyburn, *Fundamenta Mathematicae*, vol. 16 (1930), p. 305.
10. G. T. Whyburn, *Bulletin of the American Mathematical Society*, vol. 44 (1938), p. 414.
11. W. T. Puckett, *American Journal of Mathematics*, vol. 61 (1939), p. 750.

TWISTED CUBICS ASSOCIATED WITH A SPACE CURVE.*†

By LOUIS GREEN.

1. Introduction. Various methods have been employed in investigating the projective differential properties of a curve immersed in ordinary space, each method having certain advantages. The procedure used here is to start with a pair of dual differential equations, to introduce certain transformations of coördinates in order to obtain canonical power-series expansions for the curve considered, and to base the remainder of the paper on these expansions. The objectives of the paper are to characterize certain configurations associated with a curve, particularly the five-point twisted cubics, and to begin the problem of interpreting a duality formula in geometrical language.

2. Analytic basis. The differential equations of a twisted curve Γ , not belonging to a linear complex, may be written in the form

$$(1.1) \quad x^{iv} + ax'' + (a' - \theta)x' + cx = 0, \quad (\theta = \text{const.} \neq 0).$$

$$(1.2) \quad \xi^{iv} + a\xi'' + (a' + \theta)\xi' + c\xi = 0,$$

ξ represents the osculating plane of Γ at the point x ; differentiation is taken with respect to a properly chosen parameter u ; and a, c are scalar functions of u . The value of θ can be chosen arbitrarily ($\neq 0$); if $\theta = -1$, these equations are the ones derived by Fubini and Čech;¹ if $\theta = -4$, then (1.1) is the canonical form of Halphen.

When u is fixed at a suitable value u_0 , a point $O (= x)$ on Γ is obtained, and a local tetrahedron of reference $D_1\{x, x', x'', x'''\}$ is formed, with a unit point chosen so that any point whose coördinates in the original system are

$$x_1x + x_2x' + x_3x'' + x_4x'''$$

will have local coördinates proportional to x_1, \dots, x_4 . It follows readily that the local coördinates x_i of a point P on Γ "sufficiently near" O are

$$x_i = \sum_{n=0}^{\infty} A_{in} \Delta u^n / n! \quad (i = 1, \dots, 4),$$

where

$$\begin{aligned} A_{10} &= 1, \quad A_{20} = A_{30} = A_{40} = 0, \\ A_{1,n+1} &= A'_{1n} - cA_{4n}, \\ A_{2,n+1} &= A_{1n} + A'_{2n} + (\theta - a')A_{4n}, \\ A_{3,n+1} &= A_{2n} + A'_{3n} - aA_{4n}, \\ A_{4,n+1} &= A_{3n} + A'_{4n}. \end{aligned} \quad (n \geq 0)$$

* Received February 13, 1939.

† Presented to the Society, September 6, 1938.

¹ *Introduction à la Géométrie Projective Différentielle des Surfaces*, 1931, p. 26.

Halphen's local tetrahedron H_1 which will be used throughout this paper can be obtained directly from D_1 . If local coördinates referred to H_1 are denoted by y_i , then the following relations hold:

$$\begin{aligned}
 (2.1) \quad & y_1/\tau = x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \alpha_{14}x_4, \\
 & y_2/\tau = \alpha_{22}x_2 + \alpha_{23}x_3 + \alpha_{24}x_4, \\
 & y_3/\tau = \alpha_{33}x_3 + \alpha_{34}x_4, \\
 & y_4/\tau = \alpha_{44}x_4, \\
 & \alpha_{12} = \phi/20\theta, \quad \alpha_{13} = (\phi^2 - 180a\theta^2)/600\theta^2, \\
 & \alpha_{14} = (\phi^3 - 1260a\phi\theta^2 - 10800a'\theta^3 + 14400\theta^4)/36000\theta^3, \\
 & \alpha_{22} = \psi, \quad \alpha_{23} = \phi\psi/15\theta, \quad \alpha_{24} = (\phi^2 - 420a\theta^2)\psi/600\theta^2, \\
 & \alpha_{33} = 2\psi^2, \quad \alpha_{34} = \phi\psi^2/10\theta, \quad \alpha_{44} = 6\psi^3, \\
 & \phi = 100c - 9a^2 - 30a'', \quad \psi = (\theta/60q_6)^{1/3}, \quad q_6 \neq 0.
 \end{aligned}$$

If non-homogeneous coördinates x, y, z are defined in the customary way,

$$x = y_2/y_1, \quad y = y_3/y_1, \quad z = y_4/y_1,$$

the equations of Γ , relative to O as origin, are found to be

$$(3.1) \quad y = x^2 + \sum_7 p_n x^n, \quad z = x^3 + \sum_6 q_n x^n.$$

The coefficients p_n, q_n ($n \geq 7$) are expressible in terms of q_6 and the coefficients in the differential equation (1.1), and are understood, of course, to be evaluated at $u = u_0$. The value of q_6 , as is the case with θ , can be chosen arbitrarily ($\neq 0$), but since a numerical choice for q_6 would prevent us from displaying the weights² of the coefficients, we merely specify that q_6 be independent of the parameter u . The values of q_7, q_8, p_7 are found from (2.1) to be

$$\begin{aligned}
 (4.1) \quad & q_7 = \phi/4200\psi^4, \\
 & q_8 = (\phi^2 - 45\phi'\theta - 360a\theta^2)/504000\psi^5\theta, \\
 & p_7 = (-\phi^2 - 60\phi'\theta - 360a\theta^2)/1512000\psi^5\theta.
 \end{aligned}$$

Let π ($=\xi$) be the plane osculating Γ at O . Then a local plane tetrahedron of reference $D_2\{\xi, \xi', \xi'', \xi'''\}$ with local coördinates proportional to ξ_1, \dots, ξ_4 , if plane coördinates in the original system are

$$\xi_1\xi + \xi_2\xi' + \xi_3\xi'' + \xi_4\xi''',$$

is formed from the differential equation (1.2). We replace D_2 by a new local plane tetrahedron H_2 , dual to H_1 . If local plane coördinates referred to H_2 are denoted by η_i , then the relations between ξ_i and η_i are the same as those between x_i and y_i in (2.1) except for the change in sign of θ .

² The weights of p_n and q_n are $n-2$ and $n-3$ respectively.

Setting

$$\xi = \eta_2/\eta_1, \quad \eta = \eta_3/\eta_1, \quad \zeta = \eta_4/\eta_1,$$

we obtain the equations of Γ in system H_2 :

$$(3.2) \quad \eta = \xi^2 + \sum_7 \pi_n \xi^n, \quad \zeta = \xi^3 + \sum_6 \kappa_n \xi^n$$

where we choose

$$\kappa_6 = q_6.$$

The values of $\kappa_7, \kappa_8, \pi_7$ are obtained from q_7, q_8, p_7 as given in equations (4.1) by changing the sign of θ . It then follows that

$$(4.2) \quad \kappa_7 = q_7, \quad \kappa_8 = (\gamma q_6 q_8 - 18 q_6 p_7 - \gamma q_7^2)/q_6, \quad \pi_7 = (24 q_6 q_8 - 63 q_6 p_7 - 28 q_7^2)/9 q_6.$$

Returning to tetrahedron H_1 and equations (3.1) we denote homogeneous plane coördinates by $\bar{\xi}_i$ and non-homogeneous coördinates by $\bar{\eta}, \bar{\zeta}$, where

$$\bar{\xi} = \bar{\xi}_3/\bar{\xi}_4, \quad \bar{\eta} = \bar{\xi}_2/\bar{\xi}_4, \quad \bar{\zeta} = \bar{\xi}_1/\bar{\xi}_4.$$

Then the plane equations of Γ in system H_1 are

$$(5) \quad \begin{aligned} \bar{\eta} &= \bar{\xi}^2/3 + 2 q_6 \bar{\xi}^5/81 - \gamma q_7 \bar{\xi}^6/\gamma 29 + (8 q_8 - 21 p_7) \bar{\xi}^7/2187 + \dots, \\ \bar{\zeta} &= \bar{\xi}^3/27 + 5 q_6 \bar{\xi}^6/\gamma 29 - 2 q_7 \bar{\xi}^7/\gamma 29 + (\gamma q_8 - 18 p_7) \bar{\xi}^8/6561 + \dots \end{aligned}$$

The transformation

$$(6) \quad \begin{aligned} \bar{\xi}_1 &= 2 \gamma q_6^3 \eta_4, \\ \bar{\xi}_2 &= 81 q_6^3 \eta_3 - 189 q_6^2 q_7 \eta_4, \\ \bar{\xi}_3 &= 81 q_6^3 \eta_2 - 378 q_6^2 q_7 \eta_3 + 441 q_6 q_7^2 \eta_4, \\ \bar{\xi}_4 &= 2 \gamma q_6^3 \eta_1 - 189 q_6^2 q_7 \eta_2 + 441 q_6 q_7^2 \eta_3 + (54 q_6^4 - 343 q_7^3) \eta_4 \end{aligned}$$

carries (5) into (3.2) and hence is the transformation from H_1 to H_2 . The coördinates of the vertices of tetrahedron H_2 referred to system H_1 are thus found to be

$$(7) \quad \begin{aligned} &(1, 0, 0, 0), \quad (\gamma q_7, q_6, 0, 0), \\ &(49 q_7^2, 14 q_6 q_7, 3 q_6^2, 0), \quad (343 q_7^3 - 54 q_6^4, 14 \gamma q_6 q_7^2, 63 q_6^2 q_7, 2 \gamma q_6^3). \end{aligned}$$

The u -derivatives of the local point coördinates x_i of system D_1 are obtained in the following way. In the original coördinate system any point z in space has coördinates

$$z = x_1 x + x_2 x' + x_3 x'' + x_4 x'''.$$

Hence, from (1.1),

$$\begin{aligned} z' &= (x'_1 - c x_4) x + (x'_2 + x_1 + \theta x_4 - a' x_4) x' \\ &\quad + (x'_3 + x_2 - a x_4) x'' + (x'_4 + x_3) x''', \end{aligned}$$

and placing $z'_i = 0$ ($i = 1, \dots, 4$), we get

$$(8) \quad \begin{aligned} x'_1 &= cx_4, & x'_2 &= -x_1 + (a' - \theta)x_4, \\ x'_3 &= -x_2 + ax_4, & x'_4 &= -x_3. \end{aligned}$$

The formulas for the derivatives of local coördinates y_i of system H_1 are found by differentiating equations (2.1), in which we choose

$$\tau = \exp(\int \alpha_{12} du),$$

and by replacing the x'_i and the x_i by their values in terms of y_i as obtained from (8) and the inverse of (2.1). The resulting equations are

$$(9) \quad \begin{aligned} \sigma y'_1 &= -63p_7y_2 - 36q_6^2y_3, \\ \sigma y'_2 &= -3q_6y_1 + 7q_7y_2 - 42p_7y_3 - 18q_6^2y_4, \\ \sigma y'_3 &= -6q_6y_2 + 14q_7y_3 - 21p_7y_4, \\ \sigma y'_4 &= -9q_6y_3 + 21q_7y_4. \end{aligned} \quad (\sigma = 3q_6/\psi)$$

Now $x = \psi \Delta u + \dots$, so that at $u = u_0$,

$$y'_i = dy_i/du = \psi dy_i/dx.$$

Hence the x -derivatives of the local coördinates y_i are gotten immediately. These have also been obtained by Miss Newton.³

3. Duality. The differential equations of Γ show that relative to Γ at 0 the dual of a point $x_i = f_i(a, c, \theta)$ in system D_1 is the plane $\xi_i = f_i(a, c, -\theta)$ in system D_2 . From the similarity of the canonical expansions (3.1) and (3.2) it follows that the dual of a point $y_i = f_i(p_n, q_m)$ referred to H_1 is the plane $\eta_i = f_i(\pi_n, \kappa_m)$ referred to H_2 . The coördinates $\bar{\xi}_i$ of this plane referred to H_1 are then obtainable from transformation (6). The problem of finding the dual, relative to Γ at 0, of a given point is therefore solved.

But there appears to be a good deal more to the problem than this. For, the concept of duality considered here is quite different from the duality theory of projective geometry. Two coincident points, for example, may have distinct dual planes. Thus, the points

$$\begin{aligned} P(lq_6q_8 + mq_6p_7 + nq_7^2, q_6q_7, 0, 0), \\ P'(l'q_6q_8 + m'q_6p_7 + n'q_7^2, q_6q_7, 0, 0), \end{aligned} \quad (l, m, n) \neq (l', m', n')$$

in system H_1 both lie on the x -axis and for properly chosen values of l', m', n' coincide. Yet their dual planes, which can be found by the method described above, are distinct. Furthermore, in order to obtain complete generality for

³ "Consecutive covariant configurations at a point of a space curve," *Transactions of the American Mathematical Society*, vol. 36 (1934), p. 61.

both the curve Γ and the position of the point O on Γ , we do not wish to specify the relations existing, at $u = u_0$, among the coefficients in the equations (3.1). We therefore regard P as one of a three-parameter family of points generating the x -axis as l, m, n vary independently over all real numbers; i. e. we fail to consider the coincidence of P and P' unless $l, m, n = l', m', n'$ respectively.

Similarly, the point

$$(kq_7, q_6, 0, 0)$$

referred to tetrahedron H_1 lies on the x -axis and coincides with P for proper choice of k . Yet their geometrical characterizations and their dual planes are completely unrelated, and the curves they generate as u varies (k, l, m, n remaining independent of u), are entirely different.

The problem of characterizing geometrically the dual of a given point relative to Γ at O is completely untouched by the formal, analytic solution above. A special case of this problem, for example, is to determine how the point P , given above, is related geometrically to its dual plane under the assumption that l, m, n are arbitrary numerical quantities. The simplest special case of the general problem is that of characterizing geometrically the dual of a point whose coördinates are expressible in terms of q_6 and q_7 alone. This problem is readily solved. For, we may write the coördinates of such a point as

$$y_i = f_i(q_6, q_7) \equiv z_i$$

when referred to tetrahedron H_1 ; hence the dual plane has coördinates

$$\eta_i = f_i(\kappa_6, \kappa_7) = z_i$$

when referred to H_2 . By means of transformation (6) we find that this plane has the equation

$$27q_6^3z_4y_1 + (81q_6^3z_3 - 189q_6^2q_7z_4)y_2 + (81q_6^3z_2 - 378q_6^2q_7z_3 + 441q_6q_7^2z_4)y_3 \\ + [27q_6^3z_1 - 189q_6^2q_7z_2 + 441q_6q_7^2z_3 + (54q_6^4 - 343q_7^3)z_4]y_4 = 0$$

when referred to H_1 . We have therefore proved the following result.

THEOREM 1. *The dual, relative to Γ at O , of a point whose coördinates are expressible in terms of q_6 and q_7 is the polar of the point with respect to a quadric Q having the equation*

$$(10) \quad 54q_6^3y_1y_4 + 162q_6^3y_2y_3 - 378q_6^2q_7y_3^2 - 378q_6^2q_7y_2y_4 + 882q_6q_7^2y_3y_4 \\ + (54q_6^4 - 343q_7^3)y_4^2 = 0.$$

Sannia has considered ⁴ a self-dual tetrahedron S whose vertices, referred to H_1 , are

⁴ "Nuova trattazione della geometria proiettivo-differenziale delle curve sghembe," *Annali di Matematica*, IV, vol. 1 (1924), pp. 1-18; vol. 3 (1926), pp. 1-25.

$$(11) \quad \begin{aligned} &O, T(\gamma q_7, 2q_6, 0, 0), N(49q_7^2, 28q_6q_7, 12q_6^2, 0), \\ &B(343q_7^3 - 216q_6^4, 294q_6q_7^2, 252q_6^2q_7, 216q_6^3). \end{aligned}$$

The quadric Q has three-point and three-plane contact with Γ at O , and has among its rulings the edges OT, ON, BT, BN of Sannia's tetrahedron. It is contained in a one-parameter family of quadrics having this property.

4. Fundamental tetrahedra. In system H_1 the osculating conic K_2 of Γ at O is given by

$$(12.1) \quad 4y_1y_3 - 3y_2^2 = 0 = y_4,$$

and its dual, the osculating quadric cone K'_2 , has the equation

$$(12.2) \quad y_3^2 - y_2y_4 = 0.$$

Any tetrahedron $\{OP_1P_2P_3\}$ with vertices defined in the following way will be called a fundamental tetrahedron of Γ at O . One vertex is at O , a second at an arbitrary point $P_1(3, t, 0, 0)$, $t \neq 0$, on the x -axis; a third vertex $P_2(3, 2t, t^2, 0)$ is the contact point of a tangent from P_1 to the conic K_2 , and the fourth vertex $P_3(1 - \alpha t^3, t, t^2, t^3)$, for an arbitrary value of α , lies on the contact line of the cone K'_2 with its tangent plane which passes through P_2 . The tetrahedra H_1, H_2, S are fundamental tetrahedra determined by the following values of t, α : $\infty, 0$; $3q_6/\gamma q_7, 2q_6$; $6q_6/\gamma q_7, q_6$; moreover, the dual of any fundamental tetrahedron is another fundamental tetrahedron.

Associated with the fundamental tetrahedra is a family of twisted cubics, T_α , having five-point contact with Γ at O , and expressed parametrically by the equations

$$(13) \quad y_1 = 1 - \alpha t^3, \quad y_2 = t, \quad y_3 = t^2, \quad y_4 = t^3.$$

All of these cubics belong to the same null system, lie on the cone K'_2 , and have K_2 as osculating conic at O .⁵

Each choice of the point P_1 , or of the plane OP_1P_3 , determines a subset of ∞^1 fundamental tetrahedra; these can be placed in a one-to-one correspondence with the cubics T_α by choosing the vertex P_3 as the intersection, besides O , of T_α and the plane OP_1P_3 . When this is done, the following relations exist:

The polars of the vertices of one of these tetrahedra with respect to the common null system of the cubics are the faces of the tetrahedra, the tangent to T_α at P_3 is the edge P_2P_3 , and the osculating plane to T_α at P_3 is the face $P_1P_2P_3$.⁶ As P_1 traces the x -axis, α remaining fixed, the edge P_1P_3 generates a cubic surface with the equation

⁵ Lane, *Projective Differential Geometry of Curves and Surfaces*, 1932, p. 29.

⁶ Su, "Note on the projective differential geometry of space curves," *Journal of the Chinese Mathematical Society*, vol. 2 (1937), pp. 98-137.

$$(14) \quad y_1 y_4^2 - 3y_2 y_3 y_4 + 2y_3^3 + \alpha y_4^3 = 0.$$

The dual of an arbitrary point on the cubic T_α can be shown to be the plane which osculates the cubic T_α .

$$y_1 = 1 - \alpha' \tau^3, \quad y_2 = \tau, \quad y_3 = \tau^2, \quad y_4 = \tau^3$$

for which

$$\alpha' = 2q_6 - \alpha,$$

at the point determined by

$$\tau = 3q_6 t / (7q_7 t - 3q_6).$$

THEOREM 2. *The dual of the five-point cubic T_α is another five-point cubic $T_{\alpha'}$ for which $\alpha + \alpha' = 2q_6$. The self-dual cubic is T_{q_6} , and the only points of this cubic which lie in their dual planes are the points O and B of Sannia's tetrahedron.*

The self-dual cubic T_{q_6} harmonically separates, with O , the points of any two dual five-point cubics. It is called the harmonic cubic by Fubini and Čech,⁷ and the coincidence cubic by Kanitani.⁸ Theorem 2 shows that all five-point cubics are also five-plane cubics, and since T_0 is the six-point cubic, then the six-plane cubic is T_α ($\alpha = 2q_6$).

5. The principal plane of a curve. Halphen's theorem⁹ on the principal plane at a point of a curve has been extended by Bompiani¹⁰ and dualized by Sannia.¹¹ We shall carry their results still further, basing all our calculations on tetrahedron H_1 .

Let the tangent developables of Γ and of a five-point cubic T_α be cut by an arbitrary plane OP_1P_3 ($\neq \pi$) passing through the x -axis, and let the plane curves of section be denoted by Γ' and T'_α , the latter being a cusped cubic. Then the following conclusions hold:

THEOREM 3.1. *If $\alpha \neq 4q_6$, the curves Γ' and T'_α have exactly six-point contact at O for all planes OP_1P_3 . If $\alpha = 4q_6$, these curves always have just seven-point contact, with the single exception that the plane given by*

$$(15.1) \quad 3q_6 y_3 - 5q_7 y_4 = 0$$

produces curves having eight-point contact.

⁷ *Geometria Proiettiva Differenziale*, vol. 1 (1926), p. 42.

⁸ "Sur les repères mobiles attachés à une courbe gauche," *Memoirs of the Ryojun College of Engineering*, vol. 6 (1933), p. 106.

⁹ "Sur les invariants différentiels des courbes gauches," *Journal de l'École Polytechnique*, vol. 28 (1880), p. 25.

¹⁰ "Sul contatto di due curve sghembe," *Memorie della Reale Accademia delle Scienze dell'Istituto di Bologna*, ser. 8, vol. 3 (1926), pp. 35-38.

¹¹ *Loc. cit.*

If α and the plane of section OP_1P_3 are both arbitrary, the cusp of T'_α lies on the twisted cubic T_α and is the vertex P_3 of the fundamental tetrahedron determined by T_α and OP_1P_3 . The cusp-tangent is the edge P_1P_3 of this tetrahedron.

Bompiani's osculants¹² for the curve Γ' , which has an inflexion at 0 for arbitrary plane OP_1P_3 , are obtained immediately. His fourth-order neighborhood of Γ' at 0 is the vertex P_1 of the fundamental tetrahedra determined by the plane OP_1P_3 , while his neighborhood of the fifth order is the edge OP_3 . His neighborhood of the sixth order is the point P_3 which lies on the five-point twisted cubic $T_\alpha (\alpha = 4q_6)$.¹³

We shall prove only the first part of our theorem. The tangent developable of Γ has the parametric equations

$$\begin{aligned} x &= u + v, \\ (16) \quad y &= u^2 + p_7u^7 + \cdots + v(2u + 7p_7u^6 + \cdots), \\ z &= u^3 + q_6u^6 + q_7u^7 + q_8u^8 + \cdots \\ &\quad + v(3u^2 + 6q_6u^5 + 7q_7u^6 + 8q_8u^7 + \cdots). \end{aligned}$$

It meets the plane OP_1P_3 , whose equation is

$$(17) \quad ty - z = 0,$$

in a curve Γ' , represented by (17) and

$$\begin{aligned} (18) \quad y &= -4x^3/t - 24x^4/t^2 - 156x^5/t^3 - (1072 + 128q_6t^3)x^6/t^4 \\ &\quad - (7668 + 1728q_6t^3 + 320q_7t^4)x^7/t^5 + \cdots. \end{aligned}$$

If the non-homogeneous equations of the cubic T_α are written in series form, its tangent developable is seen to have the equations

$$\begin{aligned} x &= u + v, \\ (19) \quad y &= u^2 - \alpha u^5 + 3\alpha^2 u^8 + \cdots + v(2u - 5\alpha u^4 + 24\alpha^2 u^7 + \cdots), \\ z &= u^3 - 2\alpha u^6 + 7\alpha^2 u^9 + \cdots + v(3u^2 - 12\alpha u^5 + 63\alpha^2 u^8 + \cdots). \end{aligned}$$

Its intersection with the plane OP_1P_3 is a curve T'_α whose equations are (17) and

$$\begin{aligned} (20) \quad y &= -4x^3/t - 24x^4/t^2 - 156x^5/t^3 - (1072 + 32\alpha t^3)x^6/t^4 \\ &\quad - (7668 + 480\alpha t^3)x^7/t^5 + \cdots. \end{aligned}$$

The desired results then follow from (18) and (20).

¹² "Per lo studio proiettivo-differenziale delle singolarità," *Bollettino della Unione Matematica Italiana*, vol. 5 (1926), p. 118.

¹³ Su, *loc. cit.* See also his paper, "On certain twisted cubics projectively connected with a space curve," *Journal of the Chinese Mathematical Society*, vol. 2 (1937), p. 59.

Dually we choose a point $P_1 (\neq 0)$ on the x -axis as the vertex of cones containing the curves Γ and T_a .

THEOREM 3.2. *If $\alpha \neq -2q_6$, these cones have exactly six-plane contact along π for all points P_1 . If $\alpha = -2q_6$, the cones always have just seven-plane contact, with the single exception that the point with coördinates*

$$(15.2) \quad (2q_7, q_6, 0, 0)$$

produces cones having eight-plane contact.

The next step would be to consider planes through 0 which do not contain the x -axis. Let Π be such a plane, cutting the tangent developables of Γ and T_a in cusped curves Γ'' and T_a'' .

THEOREM 4.1. *The order of contact of Γ'' and T_a'' at 0 is greater for $\alpha = 5q_6/2$ than for any other value of α , regardless of the position of the plane Π . When $\alpha = 5q_6/2$, the order of contact is increased still further if Π contains the line whose equations are*

$$(21.1) \quad q_6y_2 - 4q_7y_3 = 0 = y_4,$$

and is greatest for a uniquely determined position of Π , namely

$$(22.1) \quad 3q_6^2y_2 - 12q_6q_7y_3 + (18q_6p_7 - 7q_6q_8 + 20q_7^2)y_4 = 0.$$

We prefer to prove the dual theorem, and shall state here without proof several additional results of Theorem 4.1. The osculating cusped cubics at 0 of Γ'' and T_a'' coincide if, and only if, $\alpha = 5q_6/2$, independently of the plane Π . For fixed but arbitrary α let \tilde{T}_a'' be the osculating cusped cubic of T_a'' , and let Π vary in the bundle of planes through 0. Then the inflexion point of \tilde{T}_a'' generates the ruled surface (14) while the inflexion tangent of \tilde{T}_a'' forms a congruence with the following properties. The focal sheets comprise an algebraic surface S of the sixth order whose asymptotic curves are twisted cubics; on each line of the congruence the harmonic conjugate of the inflexion point with respect to the two focal points lies in the plane π ($y_4 = 0$); the developables of the congruence meet S in twisted cubics and meet π in a family of conics whose envelope is the conic K_2 . When, in particular, the inflexion tangent of \tilde{T}_a'' passes through a point P_2 on K_2 , then the plane Π is the face OP_2P_3 of the fundamental tetrahedron determined by T_a and P_2 , while the two focal points on this inflexion tangent coincide at the point P_3 of this tetrahedron.

To obtain the dual theorem, an arbitrary point P in π but not on the x -axis is chosen as the vertex of cones containing the curves Γ and T_a . For simplicity we cut these cones by the plane $y_3 = 0$, obtaining curves $\tilde{\Gamma}$ and \tilde{T}_a . Then,

THEOREM 4.2. $\tilde{\Gamma}$ and \tilde{T}_a have exactly six-point contact at 0 for all centers of projection P unless $\alpha = -q_6/2$. If $\alpha = -q_6/2$, they have just seven-point contact unless P lies on the line whose equations are

$$(21.2) \quad 3q_6y_2 - 2q_7y_3 = 0 = y_4.$$

In this latter event the curves $\tilde{\Gamma}$ and \tilde{T}_a have precisely eight-point contact at 0 except when P has coördinates

$$(22.2) \quad (3q_6q_8 - 4q_7^2, -2q_6q_7, -3q_6^2, 0),$$

when they have nine-point contact.

For, let Γ and T_a be projected upon the plane $y_3 = 0$ from the point P

$$(1, m, n, 0) \quad (n \neq 0).$$

It can be verified without difficulty that in the plane $y_3 = 0$ the projections of Γ and of T have the respective equations

$$\begin{aligned} z = & x^3 + 3mx^4/n + (9m^2 - 2n)x^5/n^2 + (28m^3 - 13mn + q_6n^3)x^6/n^3 \\ & + (90m^4 - 64m^2n + 5n^2 + 6q_6mn^3 + q_7n^4)x^7/n^4 \\ & + (297m^5 - 285m^3n + 51mn^2 + 27q_6m^2n^3 \\ & \quad - 5q_6n^4 + 7q_7mn^4 + q_8n^5)x^8/n^5 + \dots, \\ z = & x^3 + 3mx^4/n + (9m^2 - 2n)x^5/n^2 + (28m^3 - 13mn - 2\alpha n^3)x^6/n^3 \\ & + (90m^4 - 64m^2n + 5n^2 - 15\alpha mn^3)x^7/n^4 \\ & + (297m^5 - 285m^3n + 51mn^2 - 81\alpha m^2n^3 + 12\alpha n^4)x^8/n^5 + \dots \end{aligned}$$

The results follow immediately from these equations.

The Halphen-Bompiani theorem referred to above states:

THEOREM 5.1. *The locus of points projecting Γ and the six-point cubic T_0 into cones having at least seven-plane contact is the principal plane of Γ at 0:*

$$(23.1) \quad y_3 = 0.$$

If the center of projection lies on the line with equations

$$(24.1) \quad 2q_6y_2 + p_7y_4 = 0 = y_3,$$

these cones have at least eight-plane contact along the principal plane, while for a unique point W on this line nine-plane contact is obtained.

Theorem 3.2 shows that all points on the x -axis, except 0 possibly, must be excluded from the locus. Further examination indicates that the cones projecting Γ and T_0 from 0 have but five-plane contact along π , so that 0 must also be excluded.

The dual of this theorem is

THEOREM 5.2. All planes¹⁴ passing through the principal point of Γ at 0:

$$(23.2) \quad (7q_7, q_6, 0, 0),$$

intersect the tangent developables of Γ and of the six-plane cubic $T_a (\alpha = 2q_6)$ in curves having at least seven-point contact. No other planes possess this property. If the plane of section contains the line whose equations are

$$(24.2) \quad 2q_6^2y_1 - 14q_6q_7y_2 + (21q_6p_7 - 8q_6q_8 + 42q_7^2)y_3 = 0 = y_4,$$

the curves have at least eight-point contact at the principal point, while for a unique plane Ω through this line nine-point contact is obtained.

The tetrahedra H_1 and H_2 are characterized geometrically by the following dual theorems.

THEOREM 6.1. The Halphen point H of Γ at 0 is the point of intersection, besides 0, of the six-point cubic T_0 and the principal plane of Γ . Its coördinates are

$$(25.1) \quad (0, 0, 0, 1);$$

the equation of the osculating plane Θ' to T_0 at this point is

$$(26.1) \quad y_1 = 0.$$

THEOREM 6.2. The Halphen plane Θ of Γ at 0 is that osculating plane, besides π , of the six-plane cubic $T_a (\alpha = 2q_6)$ which contains the principal point (23.2) of Γ . Its equation is

$$(25.2) \quad 27q_6^3y_1 - 189q_6^2q_7y_2 + 441q_6q_7^2y_3 + (54q_6^4 - 343q_7^3)y_4 = 0;$$

the coördinates of the contact point H' of this cubic and this plane are

$$(26.2) \quad (343q_7^3 - 54q_6^4, 147q_6q_7^2, 63q_6^2q_7, 27q_6^3).$$

Furthermore, the lines HH' and $\Theta\Theta'$ are coplanar with the edge ON of Sannia's tetrahedron,¹⁵ and together with the self-dual cubic T_{q_6} serve to characterize this tetrahedron.

The principal plane of Γ and of a five-point cubic $T_a (\alpha \neq 0)$ is their common osculating plane π , while dually the principal point of Γ and of a five-plane cubic $T_a (\alpha \neq 2q_6)$ is their common point O . Bompiani's extension of Halphen's theorem was obtained only when the principal plane of the two curves is distinct from the common osculating plane, while Theorems 3.2 and 4.2 cover the case where the two planes coincide. The complete results can thus be summarized as follows:

¹⁴ With the exception of the planes through the x -axis which must be excluded.

¹⁵ See footnote 7.

The twisted cubic T_0 determines an axis of Bompiani (24.1) and a point¹⁶ of Bompiani W in the principal plane (23.1) of Γ ; the cubic T_a ($\alpha = -q_6/2$) determines an axis of Bompiani (21.2) and a point of Bompiani (22.2) in the osculating plane π , and the cubic T_a ($\alpha = -2q_6$) determines a point of Bompiani (15.2) on the tangent to Γ at 0. Dually, the cubic T_a ($\alpha = 2q_6$) determines a ray of Bompiani (24.2) and a plane of Bompiani Ω through the principal point (23.2) of Γ ; the cubic for which $\alpha = 5q_6/2$ determines a ray of Bompiani (21.1) and a plane of Bompiani (22.1) through the point 0, and the cubic with $\alpha = 4q_6$ determines a plane of Bompiani (15.1) through the tangent to Γ at 0.

When a five-point cubic T_a is projected upon the plane $y_3 = 0$ from a point P in the plane π , the projected curve is a cusped or a nodal cubic according as P is or is not on the conic K_2 . In either case an inflexion point is obtained at 0. Now, it follows from Bompiani's work¹⁷ that a plane curve with an inflexion point sustains at this point a seven-point cusped cubic and ∞^1 eight-point cubics (two of which are nodal), but that ordinarily it possesses no eight-point cusped cubic and no nine-point osculating cubic. If the curve does happen to have a nine-point cubic, then every eight-point cubic is a nine-point cubic and there exists a unique ten-point cubic.

Theorem 4.2 therefore states that the point where the line (21.2) meets the conic K_2 , namely

$$(27) \quad (q_7^2, 2q_6q_7, 3q_6^2, 0),$$

projects Γ into a curve in the plane $y_3 = 0$ which sustains an eight-point cusped cubic, and that the point (22.2) projects Γ into a curve sustaining a ten-point cubic. The points (22.2) and (27) are not the only points in the plane π , however, with these properties. It can be shown, for example, that *the locus of all points in π projecting Γ into a curve in the plane $y_3 = 0$ which sustains a ten-point cubic is the straight line joining the points (15.2) and (22.2).*

Theorems 3.1, 4.1, and 5.2 are concerned with plane sections of the tangent developables of Γ and of T_a , and show that for properly chosen cubics T_a there are certain planes which yield curves of section having contact of higher orders than are obtained ordinarily. It is natural, then, to inquire into the nature of the curve of intersection of the tangent developables of Γ and of T_a . The results are contained in the following theorem:

¹⁶ We prefer this terminology to "principal point" since we wish to use the latter for the dual of the principal plane.

¹⁷ See footnote 12.

THEOREM 7.1. *The tangent developables of Γ and of T_α intersect in the x -axis and in a residual curve C . If $\alpha = q_6$, then C consists of a single branch having four-point contact with Γ at 0. If $\alpha \neq q_6$, then C has four branches three of which are linear, each of the three passing through 0 and having four-point contact with Γ . If $\alpha = 2q_6$, the fourth branch of C does not pass through 0 but through the principal point (23.2). This branch is linear and has for tangent and for osculating plane at the principal point the ray of Bompiani (24.2) and the plane of Bompiani Ω . If $\alpha = 5q_6/2$, the fourth branch of C has a cusp at 0 with the ray of Bompiani (21.1) as cusp-tangent and the plane of Bompiani (22.1) as osculating plane. If $\alpha = 4q_6$, the fourth branch of C has an inflexion point at 0 with the x -axis as tangent and the plane of Bompiani (15.1) as osculating plane. For all other values of α the fourth branch of C is linear, passes through 0, has two-point contact with Γ , and has π as osculating plane.*

To prove this theorem we equate the space coördinates (16) of a point on the tangent developable of Γ to the space coördinates (19) of a point on the tangent developable of T_α , after first replacing the parameters u, v in the latter set of equations by μ, ν . Elimination of ν and ν from the three equations obtained yields the equation

$$\mu = u + \sum_2^{\infty} m_k u_k,$$

where, in particular,

$$m_2[m_2^3 - 2(\alpha - q_6)] = 0.$$

Three cases thus arise:

- (a) $m_2^3 = 2(\alpha - q_6) \neq 0$;
- (b) $m_2 = 0, \quad \alpha = q_6$;
- (c) $m_2 = 0, \quad \alpha \neq q_6$.

In case (b),

$$m_3 = -3q_6^2/2q_7,$$

while in (c),

$$\begin{aligned} m_3 = 0, \quad m_4 = \alpha(2\alpha + q_6)/2(\alpha - q_6), \quad m_5 = \alpha(4\alpha - q_6)q_7/2(\alpha - q_6)^2, \\ m_6 = [\alpha(4\alpha - q_6)q_7^2 + (\alpha - q_6)\{(q_6^2 - 2\alpha q_6 - 8\alpha^2)p_7 \\ + (5\alpha^2 - 2\alpha q_6)q_8\}]/2(\alpha - q_6)^3. \end{aligned}$$

We can now express v in terms of u .

- (a) $v = m_2 u^2/2 + \dots$;
- (b) $v = -q_7 u^2/3q_6 + \dots$;
- (c) three subcases must be considered:

$$(c_1) \quad \alpha \neq 0, \alpha \neq 2q_6 : v = (\alpha - q_6)u/3(2q_6 - \alpha) \\ + (q_6 - 4\alpha)q_7u^2/9(2q_6 - \alpha)^2 + \sigma u^3 + \dots,$$

where

$$\sigma = [3(\alpha - 2q_6)(5\alpha - 2q_6)q_8 - 9(\alpha - 2q_6)(4\alpha - q_6)p_7 \\ + 7(4\alpha - q_6)q_7^2]/27(2q_6 - \alpha)^3;$$

$$(c_2) \quad \alpha = 0 : v = -u/6 + \dots;$$

$$(c_3) \quad \alpha = 2q_6 : v = q_6/7q_7 + (21q_6p_7 - 8q_6q_8 - 7q_7^2)u/49q_7^2 + \dots.$$

Substituting into (16) yields the equations

$$(a) \quad \begin{aligned} x &= u + m_2u^2/2 + \dots, \\ y &= u^2 + m_2u^3 + \dots, \\ z &= u^3 + 3m_2u^4/2 + \dots; \end{aligned}$$

$$(b) \quad \begin{aligned} x &= u - q_7u^2/3q_6 + \dots, \\ y &= u^2 - 2q_7u^3/3q_6 + \dots, \\ z &= u^3 - q_7u^4/q_6 + \dots; \end{aligned}$$

$$(c_1) \quad \begin{aligned} x &= (5q_6 - 2\alpha)u/3(2q_6 - \alpha) + (q_6 - 4\alpha)q_7u^2/9(2q_6 - \alpha)^2 + \sigma u^3 + \dots, \\ y &= (4q_6 - \alpha)u^2/3(2q_6 - \alpha) + 2(q_6 - 4\alpha)q_7u^3/9(2q_6 - \alpha)^2 + 2\sigma u^4 + \dots, \\ z &= q_6u^3/(2q_6 - \alpha) + (q_6 - 4\alpha)q_7u^4/3(2q_6 - \alpha)^2 + 3\sigma u^5 + \dots; \end{aligned}$$

$$(c_2) \quad \begin{aligned} x &= 5u/6 + q_7u^2/36q_6 + \dots, \\ y &= 2u^2/3 + q_7u^3/18q_6 + \dots, \\ z &= u^3/2 + q_7u^4/12q_6 + \dots; \end{aligned}$$

$$(c_3) \quad \begin{aligned} x &= q_6/7q_7 + (21q_6p_7 - 8q_6q_8 + 42q_7^2)u/49q_7^2 + \dots, \\ y &= 2q_6u/7q_7 + (42q_6p_7 - 16q_6q_8 + 35q_7^2)u^2/49q_7^2 + \dots, \\ z &= 3q_6u^2/7q_7 + (63q_6p_7 - 24q_6q_8 + 28q_7^2)u^3/49q_7^2 + \dots. \end{aligned}$$

The theorem follows readily from these equations.

The dual theorem is

THEOREM 7.2. *The planes containing both a tangent to Γ and a tangent to T_a form an axial pencil through the x -axis and a residual family C' of planes, whose edge of regression is a curve C'' . If $\alpha = q_6$, then C'' consists of a single branch having four-plane contact with Γ at 0. If $\alpha \neq q_6$, then C'' has four branches three of which are linear, each of the three passing through 0 and having four-plane contact with Γ . If $\alpha = 0$, the fourth branch of C'' passes through the point of Bompiani W of Theorem 5.1, and has for tangent and for osculating plane at W the axis of Bompiani (24.1) and the principal plane (23.1). If $\alpha = -q_6/2$, the fourth branch of C'' passes through the point of Bompiani (22.2) where it has the axis of Bompiani (21.2) as tangent and the plane π as singular osculating plane. If $\alpha = -2q_6$, the fourth branch of C'' passes through the point of Bompiani (15.2) where*

it has the x -axis as tangent and the plane π as osculating plane. For all other values of α the fourth branch of C'' is linear, passes through 0 and has two-plane contact with Γ .

There are certain interesting relations between these two dual theorems, but we shall not consider them here.

6. Osculating quadrics at a point of a curve. The self-dual quadric (10) was found to have both three-point and three-plane contact with Γ at 0. Although there are ∞^5 quadrics with this property, there is no non-singular quadric having both four-point and four-plane contact with Γ at 0.

The general six-point quadric of Γ at 0 has the equation

$$(28) \quad m_1(y_1y_3 - y_2^2) + m_2(y_1y_4 - y_2y_3) + m_3(y_3^2 - y_2y_4) + m_4y_4^2 = 0,$$

where the m_i are arbitrary. If this quadric contains the five-point cubic T_a , then

$$\alpha m_1 = 0 = \alpha m_2 - m_4,$$

while if the quadric is to have seven-point contact with Γ at 0, then

$$m_2q_6 + m_4 = 0.$$

Hence we have the following theorem characterizing the cubic T_{-q_6} .

THEOREM 8.1. *There exists a one-parameter family of quadrics having at least six-point contact with Γ at 0 and containing the five-point cubic T_a ($\alpha \neq 0$). Neglecting the quadric cone K'_2 , seven-point contact is obtained if, and only if, $\alpha = -q_6$.*

The dual theorem, which will be omitted, characterizes the cubic T_a ($\alpha = 3q_6$). Another characterization of T_{-q_6} is due to Sannia.¹⁸ The ∞^2 quadrics having seven-point contact with Γ at 0 have in common just two points — 0 and the residual intersection

$$(q_6, 0, 0, 1)$$

of the line OH (25.1) with the five-point cubic T_{-q_6} . Dually, the ∞^2 quadrics having seven-plane contact with Γ at 0 having in common just two tangent planes — π and the plane

$$27q_6^3y_1 - 189q_6^2q_7y_2 + 441q_6q_7^2y_3 + (81q_6^4 - 343q_7^3)y_4 = 0,$$

which is coaxial with π and Θ (25.2) and osculates the five-plane cubic T_a ($\alpha = 3q_6$).

¹⁸ Loc. cit.

The fundamental tetrahedra are related to the seven-point quadrics according to the following theorem.

THEOREM 9.1. *Each point P_2 ($\neq 0$) on the osculating conic K_2 determines a unique seven-point quadric whose intersection with π is a conic tangent to K_2 at 0 and at P_2 . The tangent plane to the quadric at P_2 osculates the six-point cubic T_6 , the tangent plane to the quadric at 0 is the face OP_1P_3 common to all fundamental tetrahedra which have P_2 for a vertex, and the polar with respect to the quadric of the vertex P_1 of these tetrahedra is the face OP_2P_3 .*

If, in particular, the point P_2 is chosen at $(0, 0, 1, 0)$, then the quadric reduces to the cone

$$(29) \quad y = x^2$$

with vertex at the Halphen point.

The dual theorem states:

THEOREM 9.2. *Each plane OP_2P_3 ($\neq \pi$) tangent to the osculating quadric cone K'_2 determines a unique seven-plane quadric whose cone of tangents through O touches K'_2 along π and along OP_2P_3 . The contact point of the quadric with the plane OP_2P_3 lies on the six-plane cubic T_α ($\alpha = 2q_6$), the contact point of the quadric with π is the vertex P_1 common to all fundamental tetrahedra determined by the plane OP_2P_3 , and the pole with respect to the quadric of the face OP_1P_3 of these tetrahedra is the vertex P_2 .*

7. Consecutive configurations. A manifold M geometrically defined for each value of the parameter u of the curve Γ generates or envelopes another manifold M' as u varies. Several five-point twisted cubics can be characterized in this way.

As u varies, the five-point cubic T_α ($\alpha = \text{const.}$) generates a surface S_α and the osculating planes of the dual cubic $T_{\alpha'}$ ($\alpha + \alpha' = 2q_6$) envelope the dual surface $S_{\alpha'}$. The tangent planes to S_α along T_α form a developable D_α , while dually the osculating planes of $T_{\alpha'}$ are tangent to $S_{\alpha'}$ along a curve $D_{\alpha'}$. Then,

THEOREM 10.1. *D_α is the tangent developable of a twisted cubic except in the following four cases. If $\alpha = -4q_6$, D_α is a cubic cone with vertex at the point whose coördinates are*

$$(30.1) \quad (14q_7, 5q_6, 0, 0).$$

If $\alpha = -2q_6/3$, D_α is a cubic cone with vertex at a point,

$$(31.1) \quad (49q_7^2, 70q_6q_7, 75q_6^2, 0),$$

lying on the osculating conic K_2 . If $\alpha = 0$, D_α is the quadric cone (29).¹⁹ If $\alpha = 6q_6$, D_α is the osculating quadric cone K'_2 . That is, the characteristic curve at 0 determined by the set of all osculating cones K'_2 is the cubic T_α ($\alpha = 6q_6$) (and the x -axis); stated differently, the generators of K'_2 form a congruence as the parameter u varies, and the surface S_α ($\alpha = 6q_6$) is one focal surface.²⁰ The cubics T_α on S_α have an envelope other than Γ if, and only if, $\alpha = 6q_6$, the contact point on the envelope associated with 0 being at

$$(343q_7^3 - 750q_6^4, 245q_6q_7^2, 175q_6^2q_7, 125q_6^3).$$

The proof runs as follows. A point P on Γ near 0 determines a canonical tetrahedron $H_1(P)$. If local homogeneous coördinates referred to this tetrahedron are denoted by Y_i , then the equations of the cubic T_α associated with P are

$$(32) \quad Y_1 = 1 - \alpha\tau^3, \quad Y_2 = \tau, \quad Y_3 = \tau^2, \quad Y_4 = \tau^3.$$

Now let P have non-homogeneous coördinates (h, k, l) referred to tetrahedron H_1 at 0. From (9) and the equation

$$Y_i = y_i + (dy_i/dx)h + \dots,$$

we have

$$\begin{aligned} \rho Y_1 &= q_6 y_1 - (21p_7 y_2 + 12q_6^2 y_3)h + \dots, \\ \rho Y_2 &= q_6 y_2 - (q_6 y_1 - 7q_7 y_2/3 + 14p_7 y_3 + 6q_6^2 y_4)h + \dots, \\ \rho Y_3 &= q_6 y_3 - (2q_6 y_2 - 14q_7 y_3/3 + 7p_7 y_4)h + \dots, \\ \rho Y_4 &= q_6 y_4 - (3q_6 y_3 - 7q_7 y_4)h + \dots. \end{aligned}$$

Solving for y_i and using (32) we obtain

$$\begin{aligned} \lambda y_1 &= q_6(1 - \alpha\tau^3) + (21p_7\tau + 12q_6^2\tau^2)h + \dots, \\ (33) \quad \lambda y_2 &= q_6\tau + (q_6 - 7q_7\tau/3 + 14p_7\tau^2 + 6q_6^2\tau^3 - \alpha q_6\tau^3)h + \dots, \\ \lambda y_3 &= q_6\tau^2 + (2q_6\tau - 14q_7\tau^2/3 + 7p_7\tau^3)h + \dots, \\ \lambda y_4 &= q_6\tau^3 + (3q_6\tau^2 - 7q_7\tau^3)h + \dots, \end{aligned}$$

as parametric equations of the surface S_α . When $h = 0$, $\tau = t$, the tangent plane to S_α has the form

$$\begin{aligned} q_6(6q_6 - \alpha)t^3y_1 - (12q_6^2 + 3\alpha q_6 - 7\alpha q_7t)t^2y_2 + (6q_6^2 + 9\alpha q_6 - 14\alpha q_7t)ty_3 \\ - (5\alpha q_6 - 7\alpha q_7t - 6\alpha q_6^2t^3 + \alpha^2q_6t^3)y_4 = 0. \end{aligned}$$

¹⁹ Newton, *loc. cit.* Also Tsuboko, "On the locus of the space cubics osculating a space curve," *Memoirs of the Ryojun College of Engineering*, vol. 10 (1937), pp. 63-74.

²⁰ Wilczynski, "General projective theory of space curves," *Transactions of the American Mathematical Society*, vol. 6 (1905), p. 109. This result has also been obtained independently by Kanitani and Newton, *loc. cit.* This cubic has been called the torsal cubic of Γ at 0 by Wilczynski.

As t varies, this plane envelopes the developable surface D_a whose edge of regression has the parametric equations

$$\begin{aligned} y_1 &= -45\alpha q_6^3(2q_6 + 3\alpha)(4q_6 + \alpha) + 315\alpha^2 q_6^2(2q_6 + \alpha)q_7t - 1470\alpha^3 q_6 q_7^2 t^2 \\ &\quad + [686\alpha^3 q_7^3 + 9\alpha q_6^3(2q_6 + 3\alpha)(4q_6 + \alpha)(6q_6 - \alpha)]t^3, \\ y_2 &= -45\alpha q_6^3(2q_6 + 3\alpha)(6q_6 - \alpha)t + 210\alpha^2 q_6^2(6q_6 - \alpha)q_7 t^2 \\ &\quad - 98\alpha^2 q_6(6q_6 - \alpha)q_7^2 t^3, \\ y_3 &= -45\alpha q_6^3(4q_6 + \alpha)(6q_6 - \alpha)t^2 + 21\alpha q_6^2(4q_6 + \alpha)(6q_6 - \alpha)q_7 t^3, \\ y_4 &= -9q_6^3(2q_6 + 3\alpha)(4q_6 + \alpha)(6q_6 - \alpha)t^3. \end{aligned}$$

These equations yield all but the last statement of the theorem. To complete the proof we set

$$y_1 = 1 - \alpha t^3, \quad y_2 = t, \quad y^3 = t^2, \quad y_4 = t^3$$

in (33) and find

$$\begin{aligned} t &= \tau + (1 - 7q_7\tau/3q_6 - 7p_7\tau^2/q_6)h + \dots, \\ t &= \tau + (1 - 7q_7\tau/3q_6 - 7p_7\tau^2/q_6 - 6q_6\tau^3 + \alpha\tau^3)h + \dots, \\ t &= \tau + (1 - 7q_7\tau/3q_6 - 7p_7\tau^2/q_6 - 6q_6\tau^3 - 2\alpha\tau^3 + 7\alpha q_7\tau^4/3q_6 \\ &\quad - 14\alpha p_7\tau^5/q_6 - 6\alpha q_6\tau^6 + \alpha^2\tau^6)h/(1 + 2\alpha\tau^3) + \dots. \end{aligned}$$

The desired result then follows immediately.

The dual theorem states:

THEOREM 10.2. *The curve D_a is a twisted cubic except in the following four cases. If $\alpha' = 6q_6$, D_a is a curve of class three lying in the plane dual to the point (30.1). If $\alpha' = 8q_6/3$, D_a is a curve of class three lying in the plane dual to the point (31.1). If $\alpha' = 2q_6$, D_a is a conic lying in the Halphen plane (25.2). If $\alpha' = -4q_6$, D_a is the osculating conic K_2 ; that is, the surface S_a is the locus of all osculating conics K_2 ; ²¹ stated differently, the tangents to K_2 form a congruence as the parameter u varies, and the surface S_a ($\alpha' = -4q_6$) is one focal surface.*

8. Projections of a space curve. When the space curve Γ and one of its five-point twisted cubics T_a are projected from an arbitrary point $P(h, k, l)$, $l \neq 0$, upon the osculating plane π at 0, the projected curves Γ' , T'_a possess certain interesting relations.

The equations of Γ' are readily found to be

$$(34) \quad \begin{aligned} y &= x^2 - kx^3/l + 2hx^4/l - (3hk + l)x^5/l^2 \\ &\quad + (7h^2 + 2k - q_6kl)x^6/l^2 + \dots, \quad z = 0, \end{aligned}$$

while those of T'_a are

²¹ Tsuboko, "On the locus of the conics osculating a space curve," *Memoirs of the Ryojun College of Engineering*, vol. 10 (1937), pp. 11-17.

$$(35) \quad y = x^2 - kx^3/l + 2hx^4/l - (3hk + l + \alpha l^2)x^5/l^2 \\ + (7h^2 + 2k + 2\alpha kl)x^6/l^2 + \dots, \quad z = 0.$$

These curves always have at least five-point contact, and hence have at 0 a common osculating conic K . Higher-order contact is obtained only when $\alpha = 0$, as seen from (34) and (35) or from the theorem that the principal plane of Γ and T_α ($\alpha \neq 0$) is the osculating plane π .

The equations of K are

$$(36) \quad y = x^2 - kxy/l + (2hl - k^2)y^2/l^2, \quad z = 0.$$

This conic has six-point contact with Γ' if, and only if,

$$(37) \quad \mu_1 \equiv l^2 + 2k^3 - 3hkl = 0,$$

and has six-point contact with T'_α if, and only if,

$$(38) \quad \mu_2 \equiv \mu_1 + \alpha l^3 = 0,$$

i.e. if, and only if, the center of projection P lies on the cubic surface (14).

If P traces a line L through 0, the conic K remains unchanged. Moreover, we can readily prove

THEOREM 11. *The conics K and K_2 have double contact if, and only if, the center of projection P lies on the quadric cone K'_2 . If this is the case, the line OP and the chord of double contact of the conics are edges of a common fundamental tetrahedron.*

We shall be concerned in this section with the projective normal and the flex-ray of Γ' at 0,²² and in order that these be well-defined we must have a non-composite osculating nodal cubic of Γ' at 0, which means that we must assume that $\mu_1 \neq 0$. This assumption furthermore prevents the center of projection from lying on the six-point cubic T_0 .

The osculating nodal cubic of Γ' at 0 can be shown to have the equations

$$l^3\mu_1xy + l^2v_1y^2 - l^3\mu_1x^3 + l^2(k\mu_1 - v_1)x^2y + l(k^2\mu_1 - 2hl\mu_1 + kv_1)xy^2 \\ + (\mu_1^2 + k^2v_1 - 2hlv_1)y^3 = 0 = z,$$

where μ_1 is given by (37) and

$$v_1 \equiv 8hk^2l - h^2l^2 - 2kl^2 - 5k^4 - q_6kl^3.$$

Hence the projective normal of Γ' at 0 is expressed by

$$(39) \quad l\mu_1x + v_1y = 0 = z,$$

while the flex-ray is seen to be

²²The flex-ray of Γ' at 0 is defined as the line of inflexions of the osculating nodal cubic of Γ' at 0.

$$(40) \quad l^2\mu_1^2 + l\mu_1(2\nu_1 + k\mu_1)x + (k^2\mu_1^2 - 2hl\mu_1^2 + \nu_1^2)y = 0 = z.$$

In the same way the projective normal at 0 of T'_a is of the form

$$(41) \quad l\mu_2x + \nu_2y = 0 = z,$$

where μ_2 is given by (38), and

$$\nu_2 \equiv 8hk^2l - h^2l^2 - 2kl^2 - 5k^4 - 2\alpha kl^3,$$

while the flex-ray of T'_a at 0 has the equations

$$(42) \quad l^2\mu_2^2 + l\mu_2(k\mu_2 + 2\nu_2)x + (k^2\mu_2^2 - 2hl\mu_2^2 + \nu_2^2)y = 0 = z.$$

THEOREM 12. *As the center of projection P traces a line L through 0, the projective normal of Γ' varies in the pencil at 0 unless L is a generator of the quartic cone*

$$(43) \quad (y^2 - xz)^2 + q_6yz^3 = 0.$$

Neglecting the x -axis which must be excluded, this cone meets the quadric cone K'_2 in the z -axis. As P traces the z -axis, the projective normal of Γ' at 0 remains coincident with the y -axis.

As P traces a line L through 0, the projective normal of T'_a varies unless L is a generator of K'_2 . As P traces a generator OP_3 of K'_2 , the projective normal of T'_a at 0 coincides with the edge OP_2 of the fundamental tetrahedra determined by OP_3 .

The proof offers no difficulties. In (39) and in (41) we replace (h, k, l) by homogeneous coördinates (h_1, \dots, h_4) and demand that the resulting equation be independent of h_1 . In both cases we obtain

$$h_4x - 2h_3y = 0 = z,$$

so that

$$2lh_3\mu_i + h_4\nu_i = 0 \quad (i = 1, 2).$$

Replacing μ_i and ν_i by their values we have the results immediately.

The locus of all centers of projection P determining curves Γ' which have at 0 a fixed projective normal, say

$$(44) \quad x + my = 0 = z,$$

is a fourth-order surface S^*_Γ determined by the condition

$$(45) \quad l\mu_1m - \nu_1 = 0.$$

The only five-point twisted cubic T_a which lies on this surface is the one for which $\alpha = q_6/2$, this situation occurring only when the given projective normal of Γ' at 0 is the y -axis.

If the projective normal of T'_a is chosen as the same line (44), then P generates another fourth-order surface S^*_a determined by the condition

$$(46) \quad l\mu_2m - v_2 = 0.$$

The surfaces S^*_Γ and S^*_a coincide if, and only if, both $\alpha = q_6/2$ and (44) is the y -axis. If $\alpha = q_6/2$ but (44) is not the y -axis, these surfaces meet only in the plane π so that there is no center of projection yielding coincident projective normals at 0 for Γ' and T'_a . If $\alpha \neq q_6/2$ and (44) is the y -axis, the surfaces meet in the z -axis. In all other cases the surfaces intersect in a non-composite conic.

From (45) and (46) we obtain

$$(47) \quad l = (q_6 - 2\alpha)k/\alpha m,$$

and upon substituting into (45) we find as the equations of the conic

$$\begin{aligned} (q_6 - 2\alpha)y - \alpha mz &= 0, \\ q_6(q_6 - 2\alpha)^3 mz + (q_6 - 2\alpha)^4 x^2 - \alpha m^2 (q_6 - 2\alpha)^2 (3q_6 + 2\alpha)xz \\ &\quad + \alpha m [\alpha^2 m^3 (2q_6 + \alpha) + q_6 (q_6 - 2\alpha)^3] z^2 = 0. \end{aligned}$$

As m varies, α remaining fixed, these conics generate a surface whose equation is

$$(48) \quad q_6 y z^2 + \alpha x^2 z^2 - (3q_6 + 2\alpha) x y^2 z + (2q_6 + \alpha) y^4 + \alpha q_6 y z^3 = 0 \quad (\alpha \neq 0).$$

This surface is composite if $\alpha = -2q_6$, and degenerates when $\alpha = 0$ into the plane $y = 0$, as is evident from (47).

The flex-ray of Γ' at 0 is tangent to the osculating conic K of Γ' at the point of intersection of flex-ray and projective normal. A similar statement holds for T'_a . Hence, as P traces a line L through 0 the envelope of the flex-rays at 0 of the curves Γ' is the conic K unless L lies on the cone (43). The following theorem can be readily proved.

THEOREM 13. *Let the center of projection P trace a five-point cubic T_a ($\alpha \neq 0$). If $\alpha = 2q_6/3$, the flex-rays at 0 of the curves Γ' form a pencil through the point $(0, 1, 0, 0)$. If $\alpha = q_6/3$, or if $\alpha = q_6$, the flex-rays all pass through the point $(0, 0, 1, 0)$. In all other cases the flex-rays at 0 of Γ' envelope a non-degenerate conic. This conic has three-point contact with K_2 at 0 if, and only if, $\alpha = 4q_6/3$.*

If P traces a five-point cubic T_β , the flex-ray at 0 of T'_a ($\alpha \neq \beta$) envelopes the conic K_2 for all values of α, β .

Let the flex-ray at 0 of Γ' be a given line, say

$$(49) \quad rx + sy + 1 = 0 = z.$$

Then from (40) we find that the center of projection P lies on a space curve C_Γ which is the intersection of a quadric cone

$$(50) \quad 8hl - 5k^2 + 2rkl - (r^2 - 4s)l^2 = 0$$

and a quadric surface

$$\omega + 50q_6kl = 0,$$

where

$$\begin{aligned} \omega = & 75k + 25rl + 42h^2 + 131rhh + (r^2 - 4s)hl + (12r^2 + 160s)k^2 \\ & - (26r^3 - 104rs)kl. \end{aligned}$$

Hence C_Γ has a node at 0 with the x -axis as one tangent. If the residual tangent at 0 lies on the cone K'_2 , then the flex-ray (49) is the edge P_1P_2 of the fundamental tetrahedra determined by the residual tangent, and conversely. Ordinarily C_Γ is a quartic curve, but under the condition

$$(r^2 - 3s)^2 - 12q_6r = 0$$

it consists of a generator of the cone (43) and a twisted cubic.

If the flex-ray at 0 of T'_a is chosen as the same line (49), then the center of projection P lies on a curve C_a which is the intersection of the cone (50) and a quadric surface

$$\omega + 75\alpha kl + 25\alpha rl^2 = 0.$$

The curves C_Γ and C_a coincide if, and only if, both $\alpha = 2q_6/3$ and the flex-ray (49) passes through the point $(0, 1, 0, 0)$.

If $\alpha = 2q_6/3$ but the flex-ray does not pass through $(0, 1, 0, 0)$, there are no centers of projection P yielding coincident flex-rays at 0 for Γ' and T'_a . If $\alpha \neq 2q_6/3$ and the flex-ray is the line $y_1 = y_4 = 0$, then the locus of P is the z -axis. If $\alpha \neq 2q_6/3$ and the flex-ray passes through $(0, 1, 0, 0)$ but not through $(0, 0, 1, 0)$, then there are no points P . In all other cases there is a unique center of projection P which determines coincident flex-rays at 0 for Γ' and T'_a , the locus of these points P being the surface (48).

Results of interest can also be obtained by studying other elements associated with Γ' and T'_a , such as the focal point on the projective normal²³ or on the flex-ray, the Halphen point, the condition for a coincidence point at 0, etc.

INDIANA UNIVERSITY,
BLOOMINGTON, INDIANA.

²³ Tsuboko, "Sur la courbure projective d'une courbe," *Memoirs of the Ryojun College of Engineering*, Inouye Commemoration Volume (1934), pp. 59-74.

THE UNLOADING PROBLEM FOR PLANE CURVES.*

By PATRICK DU VAL.

This paper relates to an earlier one of mine in this Journal,¹ and uses the same notation. In particular, Clarendon type indicates matrices, capitals being used for rows or columns of geometrical entities and small letters for numerical matrices. (By an oversight for which I apologise, **E** was used instead of **e** in the former paper for the unit or identical matrix.) The transpose of a matrix is indicated by $\bar{}$; any inequality between matrices is understood to apply to corresponding elements, i. e., $\mathbf{h} \geq \mathbf{k}$ means $h_{\alpha\beta} \geq k_{\alpha\beta}$ for all α, β , and in particular $\mathbf{h} \geq \mathbf{0}$ means $h_{\alpha\beta} \geq 0$ for all α, β .

It is familiar that if $\mathbf{O} = (O_1 \cdots O_q)$ is a set of distinct points in a plane, and $\mathbf{h} = (h_1 \cdots h_q)$ an arbitrary row of non-negative numbers, then curves of sufficiently high order exist having multiplicity h_α in O_α ($\alpha = 1, \cdots, q$), and no other multiple point, and not all passing through any other one point; but that this is no longer true if some of the points \mathbf{O} are in the neighbourhoods of others, unless certain inequalities, which we call the consistency conditions, are satisfied by the assigned multiplicities h . In fact, since the multiplicity of a curve in any point is the sum of its multiplicities in points proximate to that,² the consistency conditions are

$$(1) \qquad \mathbf{m}\bar{\mathbf{h}} \geq \mathbf{0},$$

where \mathbf{m} is the matrix defined in the paper referred to, in which

$$\begin{aligned} m_{\alpha\beta} &= 1 && \text{if } \alpha = \beta, \\ &= -1 && \text{if } O_\beta \text{ is proximate to } O_\alpha, \\ &= 0 && \text{otherwise;} \end{aligned}$$

we shall call it the proximity matrix of the points \mathbf{O} .

It is also tolerably familiar that the conditions of having multiplicity h_α in O_α and multiplicities $h_\beta \cdots h_\epsilon$ in $O_\beta \cdots O_\epsilon$ (proximate to O_α) are formally satisfied by curves whose actual multiplicities in these points are $h_\alpha + 1, h_\beta - 1 \cdots h_\epsilon - 1$ respectively; since these conditions reduce essen-

* Received August 15, 1939.

¹ P. Du Val, *American Journal of Mathematics*, vol. 18 (1936), p. 285.

² F. Enriques and O. Chisini, *Teoria geometrica delle equazioni*, vol. 2, pp. 425-438.

tially to having at least h_a coincident intersections with every simple branch through O_a , at least $h_a + h_\beta$ with every simple branch touching $O_a O_\beta$, etc. This is the unloading (scaricamento) principle of Enriques³ in its simplest form. The alteration in the multiplicities consists of adding to the row \mathbf{h} the α -th row of the matrix \mathbf{m} ; so that generalising the process we clearly have:

(2) *The conditions of having multiplicities \mathbf{h} in points \mathbf{O} are formally satisfied by curves whose actual multiplicities there are*

$$\mathbf{h} + \mathbf{xm}, \quad \mathbf{x} \geq \mathbf{0}.$$

This we may regard as the general statement of the unloading principle.

The question arises, if we attempt to impose on a curve multiplicities not satisfying the conditions (1), what multiplicities will it in fact have? Enriques⁴ asserts that this question can always be answered by the application of the unloading principle, and of another which he calls that of smoothing or evening (scorrimento). The latter is in fact the solution of the problem, as far as it concerns a set of points consecutive on a simple branch, and with only one of the inequalities (1) unsatisfied, namely that which relates to the last point but one of the sequence. The attack on the problem in general is not given explicitly, but is illustrated by a comparatively simple example, though it seems to have been generally regarded as clear that a solution can always be arrived at by a finite number of unloadings and smoothings. What I shall now shew is that, given perfectly arbitrary proximity relations between the points, and perfectly arbitrary assigned multiplicities, \mathbf{h} , we can always find a set of numbers \mathbf{k} such that:

- (a) \mathbf{k} are consistent actual multiplicities; i. e., satisfy (1).
- (b) Curves with multiplicities \mathbf{k} formally satisfy the conditions for having the assigned multiplicities; i. e., $\mathbf{k} = \mathbf{h} + \mathbf{xm}$, $\mathbf{x} \geq \mathbf{0}$.
- (c) All curves satisfying (a), (b), are formally contained in the system that will be found.

Eliminating \mathbf{k} between the conditions (a), (b), it is clear that the inequalities (1) reduce to

$$\mathbf{m}(\tilde{\mathbf{m}}\tilde{\mathbf{x}} + \tilde{\mathbf{h}}) \geq \mathbf{0},$$

which we rewrite in the form

$$(i) \quad \mathbf{a}\tilde{\mathbf{x}} + \tilde{\mathbf{c}} \geq \mathbf{0};$$

³ F. Enriques and O. Chisini, *Ibid.*, vol. 2, pp. 425-438.

⁴ F. Enriques and O. Chisini, *Ibid.*, vol. 2, pp. 425-438.

where $\tilde{c} = m\tilde{h}$, and $a = m\tilde{m} = -n$ is the negative of the intersection matrix of the diminished neighbourhoods L of the points O , as explained in my former paper. What remains of the Condition (b) is of course just the inequalities

$$(ii) \quad x \geq 0;$$

while the Condition (c) clearly means that the solution x of the inequalities (i), (ii) which we seek is such that if x' is any other solution,

$$x' \geq x.$$

The matrix a is symmetrical and positive definite, and being the negative of an intersection matrix of distinct irreducible curves, has no positive element off the diagonal. It is of unit determinant, and thus a^{-1} also consists of integers; we prove first of all that $a^{-1} \geq 0$. An algebraic proof of this was first given by Coxeter,⁵ some years ago. I subsequently noticed that the result is equivalent (interpreting the matrix as the scalar product matrix of a set of vectors in Euclidean space) to the theorem that a spherical simplex which has no obtuse dihedral angle has no obtuse edges either; and of this it is not hard to construct an elementary trigonometrical proof. The simpler argument which I give here was suggested to me by Mahler.⁶

To say that $a^{-1} \geq 0$ is the same as to say that $a\tilde{z} \geq 0$ implies $z \geq 0$. Suppose if possible that some of the z 's are negative, whereas $a\tilde{z} \geq 0$; and let z' be the row of just those of the z 's that are < 0 , the rest being omitted, and a' the diagonal minor of a obtained by omitting the rows and columns corresponding to the columns omitted in z' . Then *a fortiori* $a'z' \geq 0$, since the terms omitted are all of the form $a_{\alpha\beta}z_{\beta}$, where $z_{\alpha} < 0$, $z_{\beta} \geq 0$, so that $\alpha \neq \beta$ and $a_{\alpha\beta} \leq 0$. Consequently $z'a'z' \leq 0$, which is impossible, since a' , being a diagonal minor of the positive definite matrix a , is itself positive definite.

We conclude that

(3) If y is chosen to satisfy the inequalities

$$y \geq 0, \quad y + c \geq 0,$$

then x given by

$$\tilde{x} = a^{-1}\tilde{y}$$

is a solution of the inequalities (i), (ii).

⁵ H. S. M. Coxeter, *Annals of Mathematics*, vol. 35 (1934), p. 601.

⁶ In conversation.

(Clearly the second of these conditions is identical with (i), and the first implies (ii)).

We next observe that if x^*_α is the least value of x_α in all solutions \mathbf{x} of (i), (ii), then the row of numbers \mathbf{x}^* is itself a solution. (The proof of this was suggested to me by Rado.⁷) For (ii) is clearly satisfied; and as regards the α -th of the inequalities (i), if \mathbf{x}' is a solution in which $x'_\alpha = x^*_\alpha$, we have

$$\sum_{\beta} a_{\alpha\beta} x^*_{\beta} \geq \sum_{\beta} a_{\alpha\beta} x'_{\beta},$$

since the α -th term is the same on both sides, and in every other term $a_{\alpha\beta} \leq 0$, $x^*_{\beta} \leq x'_{\beta}$; and hence of course

$$\begin{aligned} \sum_{\beta} a_{\alpha\beta} x^*_{\beta} + c_{\alpha} &\geq \sum_{\beta} a_{\alpha\beta} x'_{\beta} + c_{\alpha} \\ &\geq 0. \end{aligned}$$

In other words,

(4) *There exists a solution \mathbf{x}^* of the inequalities (i), (ii), such that if \mathbf{x} is any other solution of them*

$$\mathbf{x} \geq \mathbf{x}^*.$$

Putting \mathbf{x}^* for \mathbf{x} in (b), we obtain the value of \mathbf{k} satisfying (c).

An explicit formula for \mathbf{x}^* in terms of \mathbf{a}, \mathbf{c} is not easy to find. I am able only to give a method of finding it which involves a finite process of trial. For this it is convenient to drop the restriction on \mathbf{x} to be a row of integers, and consider instead a row \mathbf{z} of real numbers. The foregoing argument is practically unaltered, and we conclude that there is a minimum solution \mathbf{z}^* of the corresponding inequalities

$$(i'), (ii') \quad \mathbf{a}\tilde{\mathbf{z}} + \tilde{\mathbf{c}} \geq 0, \quad \mathbf{z} \geq 0;$$

now Erdős⁸ has remarked that for every α , either $z^*_{\alpha} = 0$ or $\sum_{\beta} a_{\alpha\beta} z^*_{\beta} = 0$; for if \mathbf{z} is a solution in which $z_{\alpha} > 0$, $\sum_{\beta} a_{\alpha\beta} z_{\beta} > 0$, then z_{α} can be diminished without destroying either of these inequalities, the rest of (ii') will be unaffected, and all the rest of (i') will be strengthened, since in each of them the coefficient of z_{α} is ≤ 0 . Thus we see that

(5) *If \mathbf{z}' is the row obtained from \mathbf{z}^* by omitting all elements that vanish, and \mathbf{a}', \mathbf{c}' are obtained from \mathbf{a}, \mathbf{c} by omitting the rows and columns corresponding to these, then*

$$\mathbf{a}'\tilde{\mathbf{z}}' + \tilde{\mathbf{c}}' = 0.$$

⁷ In conversation.

⁸ In conversation.

Another form of this result is the following:

$$(6) \quad \tilde{z}^* = -bc,$$

where \mathbf{b} is a matrix in which certain rows and the corresponding columns consist entirely of zeros, and the diagonal minor obtained by omitting these is the inverse of the corresponding minor of \mathbf{a} .

I have not been able to find any direct criterion to determine which are the vanishing z^* s. It is easily seen that if $z_a = 0$ in a solution of (i'), (ii'), then $c_a \geq 0$, $h_a \geq 0$, and $g_a \geq 0$ (where $\tilde{\mathbf{g}} = \tilde{\mathbf{m}}^{-1}\mathbf{h} = \mathbf{a}^{-1}\tilde{\mathbf{c}}$); these conditions however are only necessary and not sufficient for the vanishing of z_a . In an actual case we should first try putting $b_{a\beta} = b_{\beta a} = 0$ for all values of α satisfying these conditions, then for every combination of all but one of these values, and so on, constructing each time the matrix \mathbf{b} and the row \mathbf{z}^* in accordance with (6); the rows we obtain will all satisfy (ii'), and the first one to arise satisfying (i') also, is in fact \mathbf{z}^* . From this of course we obtain \mathbf{x}^* (which is what we really want) by the obvious relation

$$(7) \quad x_a^* \text{ is the least integer which is } \geq z_a^*.$$

THE UNIVERSITY,
MANCHESTER.

A COMPLETENESS THEOREM.*

By R. P. BOAS, JR.¹

1. Introduction. This note and the following one developed out of the problem of proving that the set of functions

$$(1.1) \quad e^{2nix}, \quad xe^{2nix} \quad (n = 0, \pm 1, \pm 2, \dots)$$

is complete in $L^2(-\pi, \pi)$. This problem is equivalent to the problem of showing that an entire function $F(z)$ of the form

$$(1.2) \quad F(z) = \int_{-\pi}^{\pi} e^{izt} f(t) dt, \quad f(t) \in L^2(-\pi, \pi),$$

is identically zero if $F(2n) = F'(2n) = 0$, $n = 0, \pm 1, \pm 2, \dots$; this theorem is easily proved.² If the original problem is generalized by replacing the multiplier x in (1.1) by a more general function $G(x)$, it is more satisfactory to attack the problem directly; some uniqueness theorems for entire functions can be obtained as corollaries. In the following note, on the other hand, it is the uniqueness theorem which is generalized; the two kinds of generalization lead in different directions, and are studied by different methods.

In this note, I shall establish the following completeness theorem, which is quite easily proved once the correct formulation has been found.

THEOREM 1. *Let $G(x) \in L^2(-\pi, \pi)$. The set of functions*

$$(1.3) \quad e^{2nix}, \quad G(x)e^{(2n+1)ix} \quad (n = 0, \pm 1, \pm 2, \dots)$$

is complete in $L^2(-\pi, \pi)$ if and only if

$$(1.4) \quad G(x + \pi) + G(x) \neq 0, \quad -\pi < x < 0,$$

*except perhaps on a set of measure zero.*³

* Received November 10, 1939.

¹ Most of the results of this note were obtained while the author was a National Research Fellow.

² It is contained in Theorem 1 of the following note ("Some uniqueness theorems for entire functions," *American Journal of Mathematics*, vol. 62 (1940), pp. 319-324).

³ Since completeness and closure in L^2 are equivalent properties, any element of L^2 can be approximated, in the metric of L^2 , by a sequence of linear combinations of the functions (1.3). If (1.4) is replaced by the stronger condition that $G(x)$ is essentially bounded and $|G(x + \pi) + G(x)| > \delta > 0$ almost everywhere, the functions (1.3) are easily shown to have the stronger property that any element of L^2 can be expanded in a series of them, the series converging in the L^2 metric (converging in the mean).

COROLLARY. *The set*

$$e^{2ni x}, \quad G(x)e^{2ni x} \quad (n = 0, \pm 1, \pm 2, \dots)$$

is complete in $L^2(-\pi, \pi)$ if and only if

$$G(x + \pi) - G(x) \neq 0, \quad -\pi < x < 0,$$

except perhaps on a set of measure zero.

This follows from the theorem if $G(x)$ is replaced by $e^{-ix}G(x)$. The corollary, with $G(x) = x$, corresponds to the original problem; the formulation with the set (1.3) is more suitable for generalization.

The general problem of determining necessary and sufficient conditions for the completeness of

$$\{e^{im_k x}, \quad G(x)e^{in_k x}\},$$

where $\{m_k\}$ and $\{n_k\}$ are mutually exclusive sequences containing all the integers between them, appears to be difficult; but the case in which $\{n_k\}$ is an arithmetic progression is easy, and not essentially different from Theorem 1 (see Theorem 5). Another generalization, in which the set of Fourier functions is broken into more than two sequences, is discussed in § 4.

By means of a theorem of R. E. A. C. Paley and N. Wiener, Theorems 1 and 5 can be transformed into uniqueness theorems for entire functions of exponential type. Let W_π be the class of entire functions of exponential type $\leq \pi$, belonging to L^2 on the real axis. I state only the theorem equivalent to Theorem 1.

THEOREM 2. *Let $g(z) \in W_\pi$; let $G(t)$ be the Fourier transform of $g(x)$. A necessary and sufficient condition that every $f(z) \in W_\pi$, satisfying*

$$(1.5) \quad f(2n) = \int_{-\infty}^{\infty} f(x)g(2n+1-x)dx = 0, \quad (n = 0, \pm 1, \pm 2, \dots),$$

is identically zero, is that

$$(1.6) \quad G(t + \pi) + G(t) \neq 0, \quad -\pi < t < 0,$$

except perhaps on a set of measure zero.

A slightly less general theorem, with a considerable formal difference from Theorem 2, can be obtained by application of a theorem of S. Bochner.

THEOREM 3. *Let Λ be a linear⁵ operator from $L^2(-\infty, \infty)$ to*

⁴ The entire function $f(z)$ is of exponential type c ($c > 0$) if $|f(z)| < Ae^{c|z|}$.

⁵ "Linear" means "additive, homogeneous, and continuous."

$L^2(-\infty, \infty)$, permutable with differentiation.⁶ A necessary and sufficient condition that every $f(z) \in W_\pi$ satisfying

$$(1.7) \quad f(2n) = \Lambda\{f(2n+1)\} = 0, \quad (n = 0, \pm 1, \pm 2, \dots),$$

is identically zero, is that

$$(1.8) \quad \int_{-\infty}^{\infty} \Lambda \left\{ \frac{\sin \pi u}{u} \right\} \cos \frac{\pi u}{2} e^{-itu} du \neq 0, \quad -\frac{\pi}{2} < t < \frac{\pi}{2},$$

except perhaps on a set of measure zero.⁷

For comparison with Theorem 2 of the following note, I mention the following special case of Theorems 2 and 3.

THEOREM 4. If $f(z) \in W_\pi$, q is a positive integer, and

$$(1.9) \quad f(2n) = f^{(q)}(2n) = 0, \quad (n = 0, \pm 1, \pm 2, \dots),$$

then $f(z) \equiv 0$.

Here $G(t)$ is equal to $(it)^q e^{-it}$ on $(-\pi, \pi)$, and vanishes outside $(-\pi, \pi)$; Λ is defined on $f(x) \in W_\pi$ by the relation $\Lambda\{f(x)\} = f^{(q)}(x-1)$.

2. Proof of the completeness theorem. Let $G(x)$ be a function of $L^2(-\pi, \pi)$, having the Fourier series

$$(2.1) \quad G(x) \sim \sum_{\nu=-\infty}^{\infty} \gamma_\nu e^{i\nu x}.$$

If B is a sequence of integers, I denote by $G_B(x)$ the function whose Fourier series is the part of the Fourier series of $G(x)$ with exponents in B ; that is,

$$(2.2) \quad G_B(x) \sim \sum_{\nu \in B} \gamma_\nu e^{i\nu x}.$$

I shall prove the following theorem, which includes Theorem 1 as a special case (when N is the set of odd integers).

THEOREM 5. Let N be an arithmetic progression with elements $a + kb$, $b \geq 0$ ($k = 0, \pm 1, \pm 2, \dots$); and let B be the set of all integers kb ($k = 0, \pm 1, \pm 2, \dots$). Let $G(x) \in L^2(-\pi, \pi)$. Then the set of functions

⁶ That is, when f and g are elements of $L^2(-\infty, \infty)$ such that $g(x) = f'(x)$, we have $[\Lambda f(x)]' = \Lambda g(x)$.

⁷ We can state a theorem, similar to Theorem 3, but entirely equivalent to Theorem 2, by introducing the space L^* whose elements are functions $f(x)$ which are Fourier transforms of elements $F(t)$ of $L(-\infty, \infty)$, with the norm $\|f\| = \int_{-\infty}^{\infty} |F(t)| dt$. Then Theorem 3 remains true if Λ is a linear operator from $L^2(-\infty, \infty)$ to L^* , permutable with differentiation.

$$(2.3) \quad e^{im_k x}, \quad G(x)e^{in_k x} \quad (m_k \in N, n_k \in N)$$

is complete in $L^2(-\pi, \pi)$ if and only if

$$(2.4) \quad G_B(x) \neq 0, \quad -\pi < x < \pi,$$

except perhaps on a set of measure zero.

To establish the sufficiency of (2.4), we have to show that if it is satisfied, if $F(x) \in L^2(-\pi, \pi)$, and if

$$(2.5) \quad \int_{-\pi}^{\pi} F(x)e^{im_k x} dx = 0, \quad (k = 0, \pm 1, \pm 2, \dots),$$

and

$$(2.6) \quad \int_{-\pi}^{\pi} F(x)G(x)e^{in_k x} dx = 0, \quad (k = 0, \pm 1, \pm 2, \dots),$$

then $F(x) = 0$ almost everywhere.

Relation (2.5) shows that $F(x)$ has a Fourier series of the form

$$(2.7) \quad F(x) \sim \sum_{k=-\infty}^{\infty} \alpha_k e^{-in_k x}.$$

Let $G(x)$ have the Fourier series (2.1), and let $G_B(x)$ be defined by (2.2). Then, from (2.6), we have

$$0 = \int_{-\pi}^{\pi} F(x)G_B(x)e^{in_k x} dx + \sum_{s \in B} \gamma_s \int_{-\pi}^{\pi} F(x)e^{i(n_k+s)x} dx.$$

Since $n_k + s \in N$ if $n_k \in N$ and $s \in B$, the series on the right is zero, by (2.5). Thus we have

$$\int_{-\pi}^{\pi} F(x)G_B(x)e^{in_k x} dx = 0, \quad (k = 0, \pm 1, \pm 2, \dots),$$

so that $F(x)G_B(x)$ has a Fourier series of the form

$$(2.8) \quad F(x)G_B(x) \sim \sum_{k=-\infty}^{\infty} \beta_k e^{-im_k x}.$$

But the Fourier series of $F(x)G_B(x)$ can be obtained by formal multiplication of the Fourier series of $F(x)$ and $G_B(x)$; consequently, by (2.7),

$$(2.9) \quad F(x)G_B(x) \sim \sum_{k=-\infty}^{\infty} \delta_k e^{-in_k x},$$

since $n_k - lb \in N$ for any integer l .

Now (2.8) and (2.9) are in contradiction unless $F(x)G_B(x) = 0$ almost everywhere; since, by (2.4), $G_B(x)$ is almost nowhere zero, $F(x)$ is almost everywhere zero. This completes the proof of the sufficiency of (2.4).

To establish the necessity of (2.4), we suppose that it is not satisfied. We may suppose that $G(x)$ is periodic with period 2π , and not almost everywhere zero, since the set (2.3) is certainly not complete if $G(x) = 0$ almost everywhere. Let E (of positive measure) be the set of zeros of $G_B(x)$ in $(-\pi, \pi)$; let $C(x)$ be the characteristic function of E . We have $b \neq 0$ (since if $b = 0$, $G_B(x) = \gamma_a e^{iax}$ and has no zeros); then $G_B(x)$ has period $2\pi/b$, and hence $C(x)$ has period $2\pi/b$.

Now let $F(x) = e^{-iax}C(x)$. Then

$$\int_{-\pi}^{\pi} F(x)G(x)e^{i(a+kb)x}dx = \int_{-\pi}^{\pi} C(x)G(x)e^{ikbx}dx = 0,$$

since $C(x)G(x) = 0$ for all x .

Thus (2.6) is satisfied. Also, since any integer m which is not in N has the form $m = a + kb + c$, $0 < c < b$, we have

$$\int_{-\pi}^{\pi} F(x)e^{im_kx}dx = \int_{-\pi}^{\pi} C(x)e^{i(kb+c)x}dx = 0,$$

since e^{icx} ($0 < c < b$) is orthogonal to every function of period $2\pi/b$.

We have therefore constructed, if (2.4) is not satisfied, a function $F(x)$ of L^2 , differing from zero on a set of positive measure, and satisfying (2.5) and (2.6). Hence (2.4) is a necessary condition for the completeness of the set (2.3).

3. Deduction of Theorems 2, 3, 4. Consider two functions $g(z), f(z)$ of W_{π} . By a theorem of Paley and Wiener,⁸

$$(3.1) \quad f(z) = \int_{-\pi}^{\pi} e^{izt}F(t)dt, \quad F(t) \in L^2(-\pi, \pi);$$

$$(3.2) \quad g(z) = \int_{-\pi}^{\pi} e^{izt}G(t)dt, \quad G(t) \in L^2(-\pi, \pi);$$

by Plancherel's theorem,

$$(3.3) \quad \int_{-\infty}^{\infty} f(x)g(u-x)dx = 2\pi \int_{-\pi}^{\pi} e^{iut}F(t)G(t)dt.$$

Thus if (1.5) is satisfied, we obtain

$$(3.4) \quad \int_{-\pi}^{\pi} e^{2int}F(t)dt = 0, \quad (n = 0, \pm 1, \pm 2, \dots);$$

$$(3.5) \quad \int_{-\pi}^{\pi} e^{(2n+1)it}F(t)G(t)dt = 0, \quad (n = 0, \pm 1, \pm 2, \dots).$$

⁸ R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, 1934, p. 13. For another proof, see M. Plancherel and G. Pólya, "Fonctions entières et intégrales de Fourier multiples," *Commentarii Mathematici Helvetici*, vol. 9 (1936-37), pp. 224-248; pp. 228 ff.

If (1.6) is also satisfied, we obtain, by Theorem 1, $F(t) = 0$ almost everywhere, and consequently $f(z) \equiv 0$.

On the other hand, if (1.6) fails, there is a function $F(t)$ of L^2 , differing from zero on a set of positive measure, and satisfying (3.4) and (3.5). The functions $f(z)$ and $g(z)$ defined by (3.1) and (3.2) then belong to W_π and satisfy (1.5). This completes the proof of Theorem 2.

If $G(t) = (it)^q$ on $(-\pi, \pi)$, the function defined by (3.3) is $2\pi f^{(q)}(z)$, and Theorem 4 follows.

We now consider Theorem 3. From a characterization of the operators Λ given by Bochner⁹ it follows in particular that if $f(z)$ is a function of W_π having the form (3.1), then

$$(3.6) \quad \Lambda\{f(z)\} = \int_{-\pi}^{\pi} e^{izt} G(t) F(t) dt,$$

where $G(t)$ is essentially bounded; conversely, any essentially bounded $G(t)$ defines, through (3.6), an operator Λ having the properties specified in Theorem 3.

We write

$$g(z) = \int_{-\pi}^{\pi} e^{izt} G(t) dt,$$

so that $g(z) \in W_\pi$. From (3.6) and (3.3) we have

$$\Lambda\{f(u)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) g(u-x) dx.$$

Theorem 3 now follows from Theorem 2 if we show that (1.6) and (1.8) are equivalent for any given $G(t)$. But we have

$$\begin{aligned} \frac{2 \sin \pi u}{u} &= \int_{-\pi}^{\pi} e^{iut} dt; \\ 2\Lambda \left\{ \frac{\sin \pi u}{u} \right\} &= \int_{-\pi}^{\pi} e^{iut} G(t) dt; \\ G(t) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \Lambda \left\{ \frac{\sin \pi u}{u} \right\} e^{-itu} du, \quad -\pi < t < \pi; \\ G(t+\pi) + G(t) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \Lambda \left\{ \frac{\sin \pi u}{u} \right\} e^{-itu} (e^{-i\pi u} + 1) du \\ &= \frac{2}{\pi} \int_{-\infty}^{\infty} \Lambda \left\{ \frac{\sin \pi u}{u} \right\} \cos \frac{\pi u}{2} e^{-isu} du, \end{aligned}$$

where $s = t + \frac{1}{2}\pi$, $-\pi < t < 0$.

⁹ S. Bochner, "Ein Satz über lineare Operationen," *Mathematische Zeitschrift*, vol. 29 (1929), pp. 737-743.

In the same way we can establish the theorem stated in footnote 7. We need for this the result that any linear operator Λ from $L^2(-\infty, \infty)$ to the space L^* (defined in footnote 7), permutable with differentiation, has the form

$$\Lambda\{f(x)\} = \int_{-\infty}^{\infty} e^{ixt} F(t) G(t) dt, \quad G(t) \in L^2(-\infty, \infty),$$

where

$$f(x) \sim \int_{-\infty}^{\infty} e^{ixt} F(t) dt, \quad F(t) \in L^2(-\infty, \infty).$$

This theorem can be established by an appropriate modification of Bochner's proof of his theorem cited in footnote 9.

4. A generalization. It is natural to generalize Theorem 5 by breaking the set of Fourier functions into three or more sequences instead of only two. The results which can be obtained in this way are sufficiently indicated by a special case. Let $F(x)$ belong to L^2 and have the Fourier series

$$\begin{aligned} F(x) &\sim \sum_{\nu=-\infty}^{\infty} a_{\nu} e^{i\nu x} \\ &= \sum a_{3\nu} e^{3\nu i x} + \sum a_{3\nu+1} e^{(3\nu+1) i x} + \sum a_{3\nu+2} e^{(3\nu+2) i x}. \end{aligned}$$

Let the functions whose Fourier series are the three sums on the right be respectively $F_0(x)$, $F_1(x)$, $F_2(x)$. Then we have the following theorem.

THEOREM 6. *If $G(x)$ and $H(x)$ belong to $L^2(-\pi, \pi)$, the set of functions*

$$(4.1) \quad e^{3n i x}, \quad G(x) e^{(3n+1) i x}, \quad H(x) e^{(3n+2) i x} \quad (n = 0, \pm 1, \pm 2, \dots)$$

is complete in $L^2(-\pi, \pi)$ if and only if

$$G_0(x)H_0(x) - G_1(x)H_2(x) \neq 0, \quad -\pi \leq x \leq \pi,$$

except perhaps on a set of measure zero.

This can be proved in the same way as Theorem 5.

DUKE UNIVERSITY.

SOME UNIQUENESS THEOREMS FOR ENTIRE FUNCTIONS.*

By R. P. BOAS, JR.¹

1. Introduction. A theorem of Valiron² states that an entire function of exponential type³ $k < \pi$, having a zero in each interval $(n, n+1)$ of the real axis ($n = 0, \pm 1, \pm 2, \dots$), is identically zero. If we let the zeros run together in pairs, and slightly weaken the hypothesis that the function is of type less than π , we are led to conjecture the truth of the following theorem.

THEOREM 1. *If $f(z)$ is an entire function of exponential type such that*

$$(1.1) \quad f(iy) = O(e^{k|y|}), \quad |y| \rightarrow \infty, \quad k < \pi;$$

and

$$(1.2) \quad f(2n) = f'(2n) = 0, \quad (n = 0, \pm 1, \pm 2, \dots),$$

then $f(z) \equiv 0$.

It is easy to prove Theorem 1 by considering the entire function $f(z) \csc^2 \frac{1}{2}\pi z$, but this attack fails for the following more general theorem.

THEOREM 2. *If $f(z)$ is an entire function of exponential type⁴ satisfying (1.1) and*

$$(1.3) \quad f(x) = O(e^{l|x|}), \quad |x| \rightarrow \infty,$$

with

$$(1.4) \quad l < \frac{\pi}{2} \cot \frac{\pi}{2} \left(1 - \frac{1}{2q-1}\right),$$

(q a positive integer), then $f(z) \equiv 0$ if

$$(1.5) \quad f(2n) = f^{(2q-1)}(2n) = 0, \quad (n = 0, \pm 1, \pm 2, \dots).$$

This theorem resembles Theorem 4 of the preceding note;⁵ the function $f(z)$ is now more general, but the condition (1.5) is more restrictive than the corresponding condition (1.9) of that note. We cannot use derivatives of even order in (1.5) as long as $l > 0$ in (1.3); this follows from Theorem 3, or more directly from the examples $f(z) = [\sin(\pi z/r)]^{r-1}$, $r = 2, 3, \dots$.

* Received November 10, 1939.

¹ This note was begun while the author was a National Research Fellow.

² G. Valiron, "Sur la formule d'interpolation de Lagrange," *Bulletin des Sciences Mathématiques* (2), vol. 49 (1925), pp. 181-192, 203-224; 213. I am not quoting the most precise form of Valiron's theorem.

³ The entire function $f(z)$ is of exponential type c ($c > 0$) if $|f(z)| < Ae^{c|z|}$.

⁴ It would be enough to suppose that $f(z)$ is of order less than 2; that $f(z)$ is of exponential type would then follow from (1.1) and (1.3) by a Phragmén-Lindelöf theorem.

⁵ "A completeness theorem," *American Journal of Mathematics*, vol. 62 (1940), pp. 312-318.

I shall establish a still more general theorem in which the $(2q-1)$ -th derivative in (1.5) is replaced by a linear differential operator; this theorem is somewhat similar to Theorem 3 of the preceding note.

Let $\phi(z)$ be regular in the rectangle $|x| < L$, $|y| < \pi$, so that

$$(1.6) \quad \phi(z) = \sum_{\nu=0}^{\infty} a_{\nu} z^{\nu}, \quad |z| < \min(L, \pi).$$

Writing D for d/dz , we can form, for any entire function $f(z)$ of exponential type c , the expression

$$(1.7) \quad \phi(D)f(z) = \sum_{\nu=0}^{\infty} a_{\nu} f^{(\nu)}(z);$$

the series is convergent (for all z) if $c < \min(L, \pi)$, and (as will be shown below) summable by the method of Mittag-Leffler⁶ in any case. With these conventions concerning $\phi(z)$, the following theorem holds.

THEOREM 3. *A necessary and sufficient condition that every entire function $f(z)$, of exponential type, satisfying*

$$(1.8) \quad f(iy) = O(e^{k|y|}), \quad |y| \rightarrow \infty, \quad k < \pi;$$

$$(1.9) \quad f(x) = O(e^{l|x|}), \quad |x| \rightarrow \infty, \quad l < L;$$

and

$$(1.10) \quad f(2n) = \phi(D)f(2n) = 0, \quad (n = 0, \pm 1, \pm 2, \dots),$$

should vanish identically, is that

$$(1.11) \quad \phi(z + \tfrac{1}{2}i\pi) - \phi(z - \tfrac{1}{2}i\pi) \neq 0, \quad |x| < L, \quad |y| < \tfrac{1}{2}\pi.$$

Theorem 2 is the special case where $\phi(z) = z^{2q-1}$. To see this, we observe that the zeros z_k of $(z + \tfrac{1}{2}i\pi)^n - (z - \tfrac{1}{2}i\pi)^n$, if n is odd, are determined by the equation

$$z_k + \tfrac{1}{2}i\pi = (z_k - \tfrac{1}{2}i\pi)e^{2k\pi i/n}, \quad (k = 0, 1, 2, \dots, n-1).$$

Then we have

$$\begin{aligned} z_k &= -\frac{i\pi}{2} \frac{1 + e^{2k\pi i/n}}{1 - e^{2k\pi i/n}} \\ &= \tfrac{1}{2}\pi \cot(k\pi/n), \end{aligned}$$

so that the z_k are real and outside $(-L, L)$ if

$$L < \frac{\pi}{2} \cot \frac{\pi}{2} \frac{n-1}{n} = \frac{\pi}{2} \cot \frac{\pi}{2} \left(1 - \frac{1}{2m+1}\right),$$

if $n = 2m + 1$.

⁶ See, e. g., P. Dienes, *The Taylor series*, 1931, p. 311.

2. Preliminary discussion. Let $f(z)$ be of exponential type and satisfy (1.8) and (1.9). By a theorem of Pólya,⁷ we can write

$$(2.1) \quad f(z) = \int_C e^{-zw} F(w) dw,$$

where C is a curve containing the "conjugate indicator-diagram" of $f(z)$ in its interior, and $F(w)$ is regular on and outside C . By (1.8), (1.9), and the convexity of the indicator-diagram, we see that we may take as C any curve outside the rectangle $|u| = l$, $|v| = k$, where $w = u + iv$; a suitable choice is the rectangle $|u| = l'$, $|v| = k'$, where $k < k' < \pi$, $l < l' < L$.

Let the function $\phi(z)$ of the theorem have the power series

$$(2.2) \quad \phi(z) = \sum_{\nu=0}^{\infty} a_{\nu} z^{\nu}.$$

If $\phi(z)$ is regular in a circle containing C in its interior, we clearly have

$$(2.3) \quad \begin{aligned} \int_C e^{zw} F(w) \phi(w) dw &= \sum_{\nu=0}^{\infty} a_{\nu} \int_C e^{zw} w^{\nu} F(w) dw \\ &= \sum_{\nu=0}^{\infty} a_{\nu} f^{(\nu)}(z) = \phi(D)f(z). \end{aligned}$$

If $\phi(z)$ is not regular in a circle containing C in its interior, but is regular in the rectangle $|u| < L$, $|v| < \pi$, the integral on the left of (2.3) still exists for all z . It is clear that if the series $\sum a_{\nu} w^{\nu}$ is uniformly summable on C by any linear summation method, the formal calculation in (2.3) will still be possible, and the result will be that the series $\sum a_{\nu} f^{(\nu)}(z)$ is summable by the same method, with the integral on the left of (2.3) as its sum. Now the power series of $\phi(z)$ is uniformly summable by any Mittag-Leffler method in any closed subset of its Mittag-Leffler star,⁸ which includes at least the rectangle $|u| < L$, $|v| < \pi$. If we take, for definiteness, the summation method defined by Lindelöf's function⁹

$$(2.4) \quad E(\alpha) = \sum_{n=0}^{\infty} \frac{\alpha^n}{\{\log(n+2)\}^n} = \sum_{n=0}^{\infty} c_n \alpha^n,$$

we shall have

$$\int_C e^{zw} F(w) \phi(w) dw = \lim_{\alpha \rightarrow \infty} \frac{1}{E(\alpha)} \sum_{n=0}^{\infty} S_n(z) c_{n+1} \alpha^{n+1},$$

where

$$S_n(z) = \sum_{\nu=0}^n a_{\nu} f^{(\nu)}(z).$$

⁷ G. Pólya, "Untersuchungen über Lücken und Singularitäten von Potenzreihen," *Mathematische Zeitschrift*, vol. 29 (1929), pp. 549-640; 580 ff.

⁸ P. Dienes, *Leçons sur les singularités des fonctions analytiques*, 1913, p. 113.

⁹ P. Dienes, *op. cit.*, *loc. cit.*

We are therefore justified in writing

$$(2.5) \quad \phi(D)f(z) = \int_C e^{zw} F(w) \phi(w) dw.$$

3. Theorem 3: sufficiency. We first prove the sufficiency of our condition (1.11). In the first place, if (1.10) is satisfied, $g(z) = f(z) \csc \frac{1}{2}\pi z$ is an entire function, obviously satisfying

$$(3.1) \quad g(iy) = O(e^{(k-\frac{1}{2}\pi)|y|}), \quad |y| \rightarrow \infty.$$

It is easy to see that

$$(3.2) \quad g(x) = O(e^{l|x|}), \quad |x| \rightarrow \infty.$$

In fact, we have,¹⁰ for each θ in $(0, 2\pi)$, different from 0 or π ,

$$f(re^{i\theta}) = O(e^{r l |\cos \theta| + r k |\sin \theta|}), \quad r \rightarrow \infty,$$

and therefore

$$\begin{aligned} g(re^{i\theta}) &= O(e^{r l |\cos \theta| - r(\frac{1}{2}\pi - k) |\sin \theta|}) \\ &= O(e^{r l |\cos \theta|}). \end{aligned}$$

Since $g(z)$ is of order one, (2.7) follows by the Phragmén-Lindelöf theorem for an angle,¹¹ applied to $g(z)e^{z^2}$.

We now define a function $\psi(w)$, regular in $|u| < L$, $|v| < \frac{1}{2}\pi$, by the relation

$$(3.3) \quad 2i\psi(z) = \phi(z + \frac{1}{2}\pi i) - \phi(z - \frac{1}{2}\pi i).$$

Let C^* be the rectangle $|u| = l'$, $|v| = k' - \frac{1}{2}\pi$; since $g(z)$ satisfies (3.1) and (3.2), we have

$$g(z) = \int_{C^*} e^{zw} \gamma(w) dw,$$

where $\gamma(w)$ is regular on and outside C^* . We define $h(z)$ by

$$(3.4) \quad h(z) = \psi(D)g(z) = \int_{C^*} e^{zw} \psi(w) \gamma(w) dw.$$

Evidently $h(z)$ is an entire function of exponential type, satisfying

$$(3.5) \quad \begin{aligned} h(iy) &= O(e^{k'|y|}), & |y| &\rightarrow \infty, \\ h(x) &= O(e^{l'|x|}), & |x| &\rightarrow \infty. \end{aligned}$$

We are going to show that

$$(3.6) \quad h(2n) = (-1)^n \phi(D)f(2n) \quad (n = 0, \pm 1, \pm 2, \dots),$$

¹⁰ E. C. Titchmarsh, *The Theory of Functions*, 1932, p. 183.

¹¹ Or by the theorem cited in footnote 10.

so that $h(2n) = 0$, $n = 0, \pm 1, \pm 2, \dots$. Then, by Carlson's theorem,¹² $h(z) \equiv 0$. But, if $\psi(w) \neq 0$ in $|u| < L$, $|v| < k$ (which is assumed in (1.11)), the function $\omega(w) = 1/\psi(w)$ is regular in the same region, and

$$\omega(D)h(z) = \int_{C^*} e^{zw} \gamma(w) dw = g(z).$$

From the representation of $\omega(D)h(z)$ as a summable infinite series, it is clear that $\omega(D)h(z) \equiv 0$ if $h(z) \equiv 0$. Thus $g(z) \equiv 0$, and consequently $f(z) = g(z) \sin \frac{1}{2}\pi z \equiv 0$, which we were to prove.

It remains to establish (3.6). In what follows, any infinite series,

$$\sum_{n=0}^{\infty} A_n,$$

is to be understood as a Mittag-Leffler sum as defined in § 2, i. e. as

$$\lim_{\alpha \rightarrow \infty} \frac{1}{E(\alpha)} \sum_{p=0}^{\infty} c_{p+1} \alpha^{p+1} \sum_{n=0}^p A_n.$$

We have

$$\phi(D)f(2n) = \sum_{\nu=0}^{\infty} a_{\nu} f^{(\nu)}(2n);$$

then, since $f(z) = g(z) \sin \frac{1}{2}\pi z$,

$$(3.7) \quad (-1)^n \phi(D)f(2n) = \sum_{\nu=0}^{\infty} a_{\nu} \sum_{\mu=0}^{\nu} \binom{\nu}{\mu} \sigma_{\nu-\mu} g^{(\mu)}(2n),$$

where

$$\begin{aligned} \sigma_m &= (-1)^n (\sin \tfrac{1}{2}\pi z)^{(m)} \Big|_{z=2n} \\ &= -\tfrac{1}{2}i[i^m - (-i)^m] (\tfrac{1}{2}\pi)^m. \end{aligned}$$

Now we have, uniformly for $w + \frac{1}{2}i\pi$ in a closed subset of the star of $\phi(w)$, and in particular for $|u| \leq l'$, $|v| \leq k' - \frac{1}{2}\pi$,

$$\begin{aligned} \phi(w + \tfrac{1}{2}i\pi) &= \sum_{\nu=0}^{\infty} a_{\nu} (w + \tfrac{1}{2}i\pi)^{\nu} \\ &= \sum_{\nu=0}^{\infty} a_{\nu} \sum_{\mu=0}^{\nu} \binom{\nu}{\mu} (\tfrac{1}{2}\pi)^{\nu-\mu} w^{\mu} i^{\nu-\mu}; \end{aligned}$$

and there is a similar expression for $\phi(w - \frac{1}{2}i\pi)$. Combining these expressions, we have (referring to (3.3))

$$\psi(w) = \sum_{\nu=0}^{\infty} a_{\nu} \sum_{\mu=0}^{\nu} \binom{\nu}{\mu} \sigma_{\nu-\mu} w^{\mu},$$

the series being uniformly summable on C^* . Consequently we may substitute this expression for $\psi(w)$ in (3.4) and integrate "termwise" along C^* , obtaining

¹² E. C. Titchmarsh, *op. cit.*, p. 186.

$$\begin{aligned} h(2n) &= \sum_{\nu=0}^{\infty} a_{\nu} \sum_{\mu=0}^{\infty} \binom{\nu}{\mu} \sigma_{\nu-\mu} g^{(\mu)}(2n) \\ &= (-1)^n \phi(D) f(2n) \end{aligned}$$

by (3.7). This is (3.6); the proof of the sufficiency part of Theorem 3 is thus complete.

4. Theorem 3: necessity. Suppose that (1.11) is not satisfied. We have to construct an entire function of exponential type satisfying (1.8) and (1.9) (with some $k < \pi$ and $l < L$), and (1.10), but not vanishing identically.

Since (1.11) is not satisfied, there is at least one point w_0 , with $|u_0| < L$, $|v_0| < \frac{1}{2}\pi$, such that

$$(4.1) \quad \phi(w_0 + \tfrac{1}{2}i\pi) - \phi(w_0 - \tfrac{1}{2}i\pi) = 0.$$

Let

$$F(w) = \frac{A(w)}{(w - w_0 - \tfrac{1}{2}i\pi)(w - w_0 + \tfrac{1}{2}i\pi)},$$

where $A(w)$ is an entire function taking the value $i\pi$ at $w = w_0 \pm \frac{1}{2}i\pi$; thus $F(w)$ has residues $+1$ and -1 at $w = w_0 \pm \frac{1}{2}i\pi$.

We take numbers k, l , such that

$$|u_0| < l < L, \quad |v_0 \pm \tfrac{1}{2}\pi| < k < \pi;$$

then the points $w_0 \pm \frac{1}{2}i\pi$ are inside the rectangle bounded by the curve C : $|u| = l, |v| = k$. If we set

$$f(z) = \frac{1}{2\pi i} \int_C e^{zw} F(w) dw,$$

it is clear that $f(z)$ satisfies (1.8) and (1.9). Moreover, we have, calculating residues,

$$f(2n) = \frac{1}{2\pi i} \int_C e^{2nw} F(w) dw = e^{2n(w_0 + \frac{1}{2}i\pi)} - e^{2n(w_0 - \frac{1}{2}i\pi)} = 0, \\ (n = 0, \pm 1, \pm 2, \dots);$$

and

$$\begin{aligned} \phi(D)f(2n) &= \frac{1}{2\pi i} \int_C e^{2nw} F(w) \phi(w) dw \\ &= e^{2n(w_0 + \frac{1}{2}i\pi)} \phi(w_0 + \tfrac{1}{2}i\pi) - e^{2n(w_0 - \frac{1}{2}i\pi)} \phi(w_0 - \tfrac{1}{2}i\pi) \\ &= (-1)^n e^{2nw_0} \{ \phi(w_0 + \tfrac{1}{2}i\pi) - \phi(w_0 - \tfrac{1}{2}i\pi) \} \\ &= 0, \end{aligned} \quad (n = 0, \pm 1, \pm 2, \dots),$$

by (4.1).

ERGODIC CURVES AND THE ERGODIC FUNCTION.*

By RICHARD KERSHNER.

1. Introduction. Let M denote a bounded subset of the Euclidean plane and let $\epsilon > 0$ be fixed. Then, following M. H. Martin¹ we have

DEFINITION 1. A continuous curve

$$(1) \quad C: x = x(t), y = y(t); \quad 0 \leq t \leq 1;$$

will be said to be ϵ -ergodic to M (or to have the property (ϵ) with respect to M ²) if, for every point of M , there is a point of C at a distance $\leq \epsilon$. In general, an arbitrary set C satisfying this last condition will be said to have the property (ϵ) with respect to M .

DEFINITION 2. A continuous rectifiable curve (1) will be called an ϵ -ergodic curve for M if it is ϵ -ergodic to M and such that its length $\Lambda(\epsilon)$ is an absolute minimum for the lengths of all continuous rectifiable curves ϵ -ergodic to M .

DEFINITION 3. The length $\Lambda(\epsilon)$ of an ϵ -ergodic curve for M , considered for varying ϵ , is called the ergodic function for M .

Martin has shown³ that, for arbitrary M and $\epsilon > 0$, there is at least one $C = C(\epsilon)$ satisfying Definition 2, so that the function $\Lambda(\epsilon)$ of Definition 3 is well defined for all $\epsilon > 0$. This function is clearly non-negative and non-increasing with ϵ . Recently⁴ Martin has shown it to be a continuous function of ϵ . He had previously pointed out⁵ that $\Lambda(\epsilon) \rightarrow \infty$ as $\epsilon \rightarrow 0$ unless M is a point set lying on a continuous rectifiable curve. In the last section of the present paper the description of the asymptotic behavior of $\Lambda(\epsilon)$, for small ϵ , is extended by showing that, for an arbitrary set M ,

$$\lim_{\epsilon \rightarrow 0} 2\epsilon\Lambda(\epsilon) = \text{meas } \bar{M},$$

* Received March 16, 1939.

¹ M. H. Martin, "Ergodic curves," *American Journal of Mathematics*, vol. 58 (1936), pp. 727-734.

² Cf. A. Errera, "Un Problème de Géométrie Infinitésimale," *Académie Royale de Belgique Mémoires*, vol. 12 (1932), p. 4.

³ *Loc. cit.*, 1, p. 731.

⁴ M. H. Martin, "Note on the continuity of the ergodic function," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 541-546.

⁵ *Loc. cit.*, 1, p. 733.

where \bar{M} is the closure of M . It should be mentioned that this last section can be read independently of what precedes it.

At the present stage it seems hopeless to expect the explicit determination of $C(\epsilon)$, for all $\epsilon > 0$, for even the simplest sets M of positive plane measure. However it is possible to find considerable information about the nature of $C(\epsilon)$ both locally and in the large. The greater part of this paper is devoted to investigations of this nature. The main results are that C has no double points, has at every point a right and left hand tangent, has a well-defined tangent up to a countable number of corners and has no cusps.

2. Preliminary lemma. This section will be devoted to a general lemma on the parametrization of rectifiable curves which will be very useful in the sequel. This lemma may also be stated in such a way as to have, apparently, nothing to do with parametrization; viz.,

LEMMA 1. *Any continuous rectifiable curve C may be uniformly approximated by simple polygons, whose lengths approximate that of C .*

Proof. The proof of this Lemma 1 is routine and will simply be outlined. First one chooses a polygon B_1 which approximates C . Then, if B_1 has multiple points which are not simple isolated double points one performs a slight deformation so as to obtain a polygon B_2 approximating B_1 and such that all multiple points are simple isolated double points (and there are only a finite number of these). Now let the polygon B_2 be traced in a definite manner and suppose that at a given double point p the polygon B_2 actually crosses itself when traced in this manner. Then if the sense of tracing is reversed along that portion of B_2 which consists of a closed curve through p , B_2 will no longer cross itself at p . Then, evidently, the double point p may be "pulled apart" without introducing any new double points. In this way the (finite number of) double points of B_2 may be removed and a simple polygon B_3 found which approximates B_2 and therefore C .

A restatement of Lemma 1 which explicitly introduces the parametric representation (1) of C will also be convenient. First

DEFINITION 4. *The parametric representation (1) of C will be said to be non-crossing if the following condition is satisfied: Let $t_1 < t_2$ be the parameters of any double point of C and let Γ be any simple closed curve containing this double point which meets the four branches of C corresponding to $t < t_1$, $t > t_1$, $t < t_2$, $t > t_2$. Let p_1, p_2, p_3, p_4 be the four points of Γ whose parameters are, respectively, the greatest, least, greatest, least value of t satisfying these four inequalities and giving points on Γ . Then p_1, p_2 do not separate p_3, p_4 on Γ .*

Clearly, if C has no double points, then any parametric representation is non-crossing. On the other hand Lemma 1 shows that

LEMMA 1 bis. *Any continuous rectifiable curve C has a parametric representation (1) which is non-crossing.*

Proof. In fact one simply chooses a sequence of simple polygons converging to C even in length, parametrizes each polygon by its arc length, and lets the limit of these parametrizations define a parametrization for C . Then it is very easily verified that Definition 4 is satisfied by this representation.

ASSUMPTION A. *In the sequel it will always be assumed that the parametric representation*

$$(1) \quad C: x = x(t), y = y(t); \quad 0 \leq t \leq 1;$$

satisfies Definition 4.

3. Terminology and notation. Let C be an ϵ -ergodic curve (1) for M . Then

DEFINITION 5. *By the point $[t]^a$ of C will be meant the point with coördinates $x(t), y(t)$.*

DEFINITION 6. *By the arc (s, u) , (where $s < u$ is always understood) will be meant the set of all points $[t]$ with $s < t < u$. If one or both of these parentheses is replaced by a square bracket, it is understood that equality is allowed at the corresponding end or ends of this last inequality.*

DEFINITION 7. *The arc (s, u) of C will be called non-significant if the point set consisting of the two arcs $[0, s]$, $[u, 1]$ of $C - (s, u)$ has the property (ϵ) with respect to M . In the contrary case (s, u) will be called significant.*

DEFINITION 8. *A point $[t]$ of C will be called non-significant if some arc (s, u) , with $s < t < u$, is non-significant. In the contrary case,⁷ that every such (s, u) is significant, the point $[t]$ will be called significant.*

DEFINITION 9. *A point of \bar{M} will be said to be salient to (s, u) if it is at a distance $\leq \epsilon$ from some point of (s, u) and at a distance $> \epsilon$ from every point of $C - (s, u)$. Such a point will be denoted by $p(s, u)$, and the totality of all such points by $P(s, u)$.*

^a The letters s, t, u will all be used as parameter values in the sequel but the generic parameter value will always be referred to as t .

⁷ Here, and throughout the paper, it is assumed that we are not dealing with an end point 0 or 1. This assumption is made simply for simplicity of statement and the definitions and results are readily extended to include the case of end points.

DEFINITION 10. A point of \bar{M} will be said to be salient to $[t]$ if it is at a distance ϵ from $[t]$ and at a distance $\geq \epsilon$ from any point of C . Such a point will be denoted by $p(t)$, and the totality of all such points by $P(t)$.

DEFINITION 11. The circle of radius ϵ about a point $p(t)$ will be called a salient circle through $[t]$ and denoted by $S(t)$.

DEFINITION 12. A significant point $[t]$ will be called simply significant if there is exactly one salient circle $S(t)$ through $[t]$.

DEFINITION 13. A significant point $[t]$ will be called doubly significant if there are exactly two salient circles $S_1(t)$ and $S_2(t)$ through $[t]$ and if these are mutually tangent.

DEFINITION 14. A significant point $[t]$ will be called multisignificant if there are at least two intersecting salient circles through $[t]$.

Notice that a salient circle $S(t)$ passes through $[t]$ but has no point of C in its interior. It will be shown, in the next section, that there is at least one $S(t)$ through every significant point $[t]$ so that Definitions 12, 13, 14 provide a classification of all significant points. Notice also that, according to Definition 10, the set $P(t)$ of points salient to $[t]$ is a closed set lying on the boundary of the ϵ -circle about $[t]$. In view of this one can make the following definition:

DEFINITION 15. Let $[t]$ be significant and let S_ϵ be the circle of radius ϵ about $[t]$. Let A_ϵ be a ⁸ closed arc of S_ϵ of minimum length which contains the set $P(t)$. Then the angular measure $\theta(t)$ of this arc is called the salient angle for $[t]$, and its endpoints are called the terminal salient points $p_1(t)$, $p_2(t)$ for $[t]$.

Notice that

$$(2a) \quad \theta(t) = 0, \text{ if } [t] \text{ is simply significant;}$$

$$(2b) \quad \theta(t) = \pi, \text{ if } [t] \text{ is doubly significant;}$$

$$(2c) \quad 0 < \theta(t) \leq \pi, \text{ if } [t] \text{ is multisignificant.}$$

The last of these relations comes from the fact that if $\theta(t) > \pi$, then the set of circles $S(t)$ would completely cover some neighborhood of $[t]$ so that $[t]$ would be an isolated point of C , contradicting the continuity of C . It will be shown in the next section that the equality sign in (2c) cannot hold.

⁸This arc will not be unique if $[t]$ is doubly significant but its length and end points are, of course, determined.

DEFINITION 16. By $K(p_1, p_2)$ and $\rho(p_1, p_2)$, where p_1 and p_2 are points, will be understood the linear segment joining p_1 and p_2 , and its length, respectively.

4. **Fundamental relations.** This section will be devoted to a study of some of the fundamental relations between the concepts introduced in the preceding section.

LEMMA 2. Suppose the point $[t]$ of C is significant. Then there is at least one salient circle $S(t)$ through $[t]$.

Proof. Let (s_i, u_i) be a sequence of arcs such that

$$(3a) \quad s_i < s_{i+1} < t < u_{i+1} < u_i; \quad (i = 1, 2, \dots)$$

and

$$(3b) \quad \lim_{i \rightarrow \infty} s_i = \lim_{i \rightarrow \infty} u_i = t.$$

Then, by Definition 8, (s_i, u_i) is significant for all i . According to Definition 7 this means that the set \bar{P}_i which is defined as the closure of the set of points $P(s_i, u_i)$, is non-empty. In fact Definition 6 implies the existence of points of M which are at a distance $> \epsilon$ from $[0, s_i] + [u_i, 1]$. These points must be at a distance $\leq \epsilon$ from (s_i, u_i) since C has the property (ϵ) with respect to M and consequently must be points of $P(s_i, u_i)$ by Definition 9.

It is clear from Definition 9 that $P(s_i, u_i) \supset P(s_{i+1}, u_{i+1})$ in view of (3a).

Thus

$$(4) \quad \bar{P}_i \supset \bar{P}_{i+1}, \quad i = 1, 2, \dots; \quad \bar{P}_i \text{ not empty.}$$

By a well-known theorem on closed sets (4) implies that there is a closed, non-empty set $\Pi = \Pi(t)$ such that

$$(5) \quad \Pi(t) = \lim_{i \rightarrow \infty} \bar{P}_i = \bigcap_{i=1}^{\infty} \bar{P}_i.$$

Let p_0 be any point of $\Pi(t)$. Then, by (5) and the definition of P_i , p_0 is at a distance $\leq \epsilon$ from (s_i, u_i) for all i . Thus, by (3b), p_0 is at a distance $\leq \epsilon$ from $[t]$. On the other hand p_0 is at a distance $\geq \epsilon$ from any point of $C - (s_i, u_i)$ for every i , and so at a distance $\geq \epsilon$ from C . Thus, by Definition 10, p_0 is a salient point $p(t)$ to $[t]$, and the ϵ -circle about p_0 is a salient circle through $[t]$. This completes the proof of Lemma 2.

Incidentally the following fact, which will be useful in the sequel, has been demonstrated during the course of the last proof.

LEMMA 3. Let $s < t < u$. Then

$$\lim_{s, u \rightarrow t} \bar{P}(s, u) \subset P(t)$$

where $\bar{P}(s, u)$ is the closure of $P(s, u)$.

Actually this was proved above only in the case that $[t]$ was a significant point, but the lemma is vacuously true if $[t]$ is non-significant.

Lemma 3 is quite restrictive in case $[t]$ is simply or doubly significant; i. e. if $P(t)$ consists of one or of two points, but in case $[t]$ is multisignificant a stronger result is needed and will be given next.

LEMMA 4. *Let $[t]$ be a multisignificant point of C and let $p_1(t), p_2(t)$ be the two terminal salient points to $[t]$. Let $s < t < u$. Then, if the notation is chosen appropriately*

$$\lim_{s \rightarrow t} \bar{P}(s, t) \subset (p_1(t)); \quad \lim_{u \rightarrow t} \bar{P}(t, u) \subset (p_2(t)).$$

Proof. First, by Definition 9,

$$\bar{P}(s, t) + \bar{P}(t, u) \subset \bar{P}(s, u).$$

Now, by Lemma 3,

$$\lim_{s, u \rightarrow t} \bar{P}(s, u) \subset P(t)$$

so that, à fortiori,

$$(6) \quad \lim_{s \rightarrow t} \bar{P}(s, t) + \lim_{u \rightarrow t} \bar{P}(t, u) \subset P(t).$$

It will now be shown that no point $p(t)$ different from $p_1(t), p_2(t)$ can be a point of the left side of (6). To this end let $p_3(t)$ be a fixed point of $P(t)$ which is not terminal. The point $p_3(t)$ is then an interior point of the arc A_ϵ mentioned in Definition 15. Let S_η be a circle with center at $p_3(t)$ and radius η , where η is so small that S_η does not cross either chord $K([t], p_1(t))$, $K([t], p_2(t))$. The circle S_η is divided by an arc of A_ϵ into two parts, one lying within and one outside S_ϵ . The points of S_η within or on S_ϵ are all at a distance $\leq \epsilon$ from $[t]$ and so, according to Definition 9, cannot belong to $P(s, t) + P(t, u)$ for any $s < t < u$. On the other hand, the points of S_η outside S_ϵ are obviously at a distance $> \epsilon$ from any point of (s, u) for sufficiently small $u - s$ in view of the fact that (s, u) cannot enter either circle $S_1(t), S_2(t)$ of radius ϵ about $p_1(t), p_2(t)$, respectively. Thus no point of S_η is a point of $P(s, t) + P(t, u)$ and $p_3(t)$ cannot be a point of

$$\lim_{s \rightarrow t} \bar{P}(s, t) + \lim_{u \rightarrow t} \bar{P}(t, u).$$

Thus (6) may be strengthened to

$$\lim_{s \rightarrow t} \bar{P}(s, t) + \lim_{u \rightarrow t} \bar{P}(t, u) \subset (p_1(t)) + (p_2(t)).$$

The separation of this last relation into the two separate inclusions required by Lemma 4 is accomplished very easily by recalling the Assumption A stating that the two arcs (s, t) and (t, u) do not cross so that one is "nearer"

$p_1(t)$ and the other "nearer" $p_2(t)$. The details of the separation will not be given as they are readily supplied.

The next two lemmas are immediate consequences of the definitions of the concepts involved and are stated simply for reference.

LEMMA 5. *The set of significant points of $C(\epsilon)$ form a closed subset of $C(\epsilon)$.*

LEMMA 6. *For any t ,*

$$\lim_{u \rightarrow t} \bar{P}(u) \subset P(t).$$

It should be mentioned that the limit here, as in Lemma 3 and Lemma 4, is the ordinary point set limit; i. e., the set of all limit points obtained using any sequence of u -values and any choice of particular points of $\bar{P}(u)$. Notice that actually $P(u) = \bar{P}(u)$; (i. e., $P(u)$ is closed) but $\bar{P}(u)$ has been written for the sake of the analogy with Lemmas 3 and 4.

LEMMA 7. *For any t_0 ,*

$$\limsup_{u \rightarrow t_0} \theta(u) \leq \theta(t_0).$$

Proof. This is an easy consequence of Lemma 6. In fact if $\limsup_{u \rightarrow t} \theta(u) = 0$, there is nothing to prove. Then suppose $\limsup_{u \rightarrow t} \theta(u) = \theta > 0$ and let $\{u_i\}$, $i = 1, 2, \dots$, be a sequence of t values such that $u_i \rightarrow t_0$ and

$$(7) \quad \lim_{i \rightarrow \infty} \theta(u_i) = \theta > 0.$$

Now let $p_1(u_i), p_2(u_i)$ be the terminal salient points for $[u_i]$. Then the points $p_1(u_i)$ are an infinite set in a bounded closed region \bar{M} and have at least one cluster point p_1 in \bar{M} . Let $\{u_{n_i}\}$ be a subsequence of the $\{u_i\}$ such that

$$(8) \quad \lim_{i \rightarrow \infty} p_1(u_{n_i}) = p_1.$$

Then, in view of Definition 15, the relations (7) and (8) imply that

$$(9) \quad \lim p_2(u_{n_i}) = p_2$$

exists; and, further, that p_1 and p_2 are two points on the ϵ -circle about $[t]$ which are separated by an angle θ on this circle. Since (8) and (9) imply that p_1 and p_2 are points of $P(t)$, in view of Lemma 6, the proof of Lemma 7 is complete.

In general the inequality given by Lemma 7 cannot be replaced by equality; but the case when this is possible, namely when $\limsup_{u \rightarrow t} \theta(u) = \pi$, deserves special mention in view of its later usefulness. In particular,

LEMMA 8. *Let $[t]$ be a limit point of doubly significant points. Then either $[t]$ is doubly significant or multisignificant with $\theta(t) = \pi$.*

As mentioned before, it will later be shown that the second alternative here cannot actually occur so this Lemma 8 will be strengthened.

5. Local properties. This section will be devoted to establishing a few local properties of C which will be needed later.

LEMMA 9. *Let $[t_0]$ be a non-significant point of C . Then if $u - s$ is sufficiently small, $s < t_0 < u$, the arc (s, u) is linear.*

Proof. According to Definition 8, some (s, u) is non-significant. Suppose this (s, u) is not linear. Then the curve obtained from C by replacing the arc (s, u) by its chord $K([s], [u])$ is a shorter curve which, according to Definition 7, has the property (ϵ) with respect to M . This contradicts the assumption that C was an ϵ -ergodic curve for M .

LEMMA 10. *Let $[t_0]$ be a simply significant point of C and let $L(t_0)$ be the line tangent to the unique salient circle $S(t_0)$ at $[t_0]$. Then if $u - s$ is sufficiently small, $s < t_0 < u$, the arc (s, u) lies in that closed half plane determined by $L(t_0)$ which does not contain $p(t_0)$.*

Proof. Suppose the statement is false; i. e., that there is a sequence of points $[t_i]$ lying in that open half plane determined by $L(t_0)$ which contains $p(t_0)$ and such that $t_i \rightarrow t_0$. For the sake of definiteness let it be supposed that $t_i < t_0$.

Let S^* be the circle of radius $\frac{1}{2}\epsilon$ about $p(t_0)$. Then, by Lemma 3, values s_0, u_0 may be chosen so that

$$\bar{P}(s_0, u_0) \subset S^*, \quad s_0 < t_0 < u_0,$$

and, à fortiori,

$$(10) \quad \bar{P}(s_0, t_0) \subset S^*.$$

Now about $p(t_0)$ draw a circle $S(t_0, \rho)$ of radius $\epsilon + \rho$, where $\rho > 0$ is so small that $S(t_0, \rho)$ intersects both arcs (s_0, t_0) and (t_0, u_0) .⁹ Let s_1 be the greatest $t < t_0$ and u_1 the least $t > t_0$ such that $[s_1]$ and $[u_1]$ are on $S(t_0, \rho)$. Then by the first paragraph of this proof there are points of (s_0, t_0) lying in that open half plane determined by $L(t_0)$ which contains $p(t_0)$. Thus some half line $K(t_0)$, terminated by $[t_0]$ and lying in this same half plane, meets (s_1, t_0) in a point $[t^*] \neq [t_0]$. Clearly $K(t_0)$ may be chosen in such a way that it does not meet S^* . It is supposed that this has been done and also that

⁹ This is possible unless one of (s_0, t_0) , (t_0, u_0) lies exactly along $S(t_0)$. In this case the argument which follows may be modified by choosing the notation so that (s_0, t_0) lies along $S(t_0)$ and then choosing $\rho = 0$, $u_1 = t_0$, $s_1 = t^* = s_0$.

$[t^*]$ is the first point distinct from $[t_0]$ where $K(t_0)$ meets (s_1, t_0) . (That there is such a first point is clear from the fact that $K(t_0)$, near $[t_0]$ lies in $S(t_0)$ and so does not meet C .)

Now, by (10),

$$P(t^*, t_0) \subset S^*.$$

This means that the only points of \bar{M} at a distance $\leq \epsilon$ from (t^*, t_0) which are not also at a distance $\leq \epsilon$ from $C - (t^*, t_0)$ are points of S^* . But, by Assumption A, the arc (t^*, t_0) , which does not cross $K(t_0)$ by the definition of t^* , cannot cross either of the two arcs comprising $(s_1, u_1) - (t^*, t_0)$. This (t^*, t_0) is separated from S^* in the convex curve $S(t_0, \rho)$ by the arc $A(s_1, u_1)$ defined by

$$A(s_1, u_1) = (s_1, t^*) + K([t^*], [t_0]) + (t_0, u_1).$$

Thus any point of S^* at a distance $\leq \epsilon$ from (t^*, t_0) is also at a distance $\leq \epsilon$ from $A(s_1, u_1)$. Then the curve

$$C - (s_1, u_1) + A(s_1, u_1) \equiv C - (t^*, t_0) + K([t^*], [t_0])$$

is ϵ -ergodic to M . Since this new curve is obviously shorter than C , this contradicts the assumption that C was an ϵ -ergodic curve for M and completes the proof of Lemma 10.

LEMMA 11. *Let $[t_0]$ be a doubly significant point of C . Then if $u - s$ is sufficiently small, $s < t_0 < u$, the arc (s, u) lies between two mutually tangent ϵ -circles through $[t_0]$.*

Proof. This is trivial (the circles being the two salient circles $S_1(t_0)$, $S_2(t_0)$ assumed in Definition 13) and has been included for reference only.

LEMMA 12. *Let $[t_0]$ be a multisignificant point of C and let $L_1(t_0)$, $L_2(t_0)$ be the tangents to the two terminal salient circles $S_1(t_0)$, $S_2(t_0)$, respectively, at $[t_0]$. Then if $u - s$ is sufficiently small, $s < t_0 < u$, the arc (s, u) is contained in that closed angle determined by $L_1(t_0)$, $L_2(t_0)$ which contains no interior point of $S_1(t_0)$ or $S_2(t_0)$.¹⁰*

Proof. The proof of Lemma 12 precisely parallels that of Lemma 10 and will not be given. It should be noticed that Lemma 4 serves here the purpose of establishing a relation like (10), where, of course, S^* will be a $\frac{1}{2}\epsilon$ -circle about $p_1(t_0)$ or $p_2(t_0)$ according to which arc is to be modified.

¹⁰ In case $\theta(t_0) = \pi$ this closed angle degenerates to a half line and is not uniquely determined by the condition that it contain no interior point of $S_1(t_0)$ or $S_2(t_0)$. However, since $[t]$ is multisignificant, there must be, in this case, a third salient circle $S_3(t_0)$ intersecting $S_1(t_0)$ and $S_2(t_0)$ (cf. definition 14). Then the half line is determined by the condition that it does not cut $S_3(t_0)$.

6. Double points. This section will be devoted to the proof of

THEOREM 1. *Let M be an arbitrary plane point set. Let $\epsilon > 0$ be fixed, and let $C = C(\epsilon)$ be an ϵ -ergodic curve for M . Then C has no double points.*

Proof. Suppose the statement is false; i. e., that there is a $t_1 \neq 0$ and a $t_2 \neq 1$,¹¹ such that $t_1 < t_2$ while $[t_1] = [t_2]$. Then there are a number of cases to be considered which are not mutually exclusive but which together exhaust all possibilities.

Case 1. The points $s_1 < t_1 < u_1$, $s_2 < t_2 < u_2$ can be chosen so that the four arcs (s_1, t_1) , (t_1, u_1) , (s_2, t_2) , (t_2, u_2) coincide (as point sets) in two identical pairs. Suppose that these four arcs have been extended so that they are as long as possible satisfying the required condition of coinciding in two pairs and such that each coincident pair have coincident end points. Then there are the following possibilities (not mutually exclusive).

Case 1.1: $s_1 = 0$. Then the curve obtained from C by deleting the arc $[s_1, t_1] = [0, t_1]$ is "shorter" in the parametric sense but identical in a point set sense with C . This contradicts the assumption that C was an ϵ -ergodic curve for M .

Case 1.2: $u_2 = 1$. A contradiction is reached, as in Case 1.1, by deleting $(t_2, u_2] = (t_2, 1]$.

Case 1.3: $s_1 \neq 0$; $u_2 \neq 1$; $[s_1] = [u_2]$. In this case $[s_1] = [u_2]$ is a double point which does not come under Case 1. For if $[s_1] = [u_2]$ came under Case 1, then the given arcs $[s_1, t_1] = [t_2, u_2]$ could be extended to longer point set identical arcs with coincident end points, contradicting the assumption that the given arcs were the longest such. Thus to exclude Case 1.3 it will be sufficient to show there are no double points which are not in Case 1.

Case 1.4: $s_1 \neq 0$; $[s_1] = [s_2]$. This is treated exactly as is Case 1.3.

Case 1.5: $s_1 \neq 0$; $[s_1] = [u_1]$. This is treated exactly as is Case 1.3.

The above five possibilities exhaust Case 1 in view of the assumption that the identical arcs had identical end points.

Case 2. The point $s_1 < t_1 < u_1$, $s_2 < t_2 < u_2$ can be chosen so that some three of the four arcs (s_1, t_1) , (t_1, u_1) , (s_2, t_2) , (t_2, u_2) are identical (as point sets). This case can be treated in essentially the same way as Case 1 and the detailed treatment will not be given. Either a contradiction is reached

¹¹ Cf. footnote 7. Of course $t_1 = 0$ and $t_2 = 1$ is not considered as a double point but merely means that C is closed, which is trivially seen to be possible.

or one is led to the existence of a double point which is not in Case 1 or Case 2. Thus it will be sufficient to prove the impossibility of this last.

Case 3. There are no salient circles through $[t_1] = [t_2]$. Then, by Lemma 2, both $[t_1]$ and $[t_2]$ are non-significant; i. e., some arcs (s_1, u_1) , (s_2, u_2) , with $s_1 < t_1 < u_1$, $s_2 < t_2 < u_2$ are non-significant. Then, by Lemma 9, these two arcs are linear. But by Assumption A these two segments through $[t_1] = [t_2]$ are non-crossing. This is possible only if $[t_1] = [t_2]$ is in Case 1.

Case 4. There is exactly one salient circle S through $[t_1] = [t_2]$. Let p be the center of S and L the tangent line to S at $[t_1] = [t_2]$. Now let s_i, u_i , $i = 1, 2$, be chosen so that

$$(11a) \quad s_1 < t_1 < u_1 < s_2 < t_2 < u_2;$$

$$(11b) \quad \bar{P}(s_i, u_i) \subset S, \quad (i = 1, 2);$$

$$(11c) \quad (s_i, u_i) \subset H, \quad (i = 1, 2);$$

where H denotes that closed half plane determined by L which does not contain p . The possibility of satisfying (11b) is assured by Lemma 3 while (11c) may be satisfied in view of Lemma 10 if $[t_i]$ is simply significant and in view of Lemma 9 if $[t_i]$ is non-significant.

Now let a circle $S(\rho)$ be drawn with center at p and with radius $\epsilon + \rho$ where $\rho > 0$ is so small that $S(\rho)$ intersects all four arcs, (s_i, t_i) , (t_i, u_i) , $i = 1, 2$. That such a $\rho > 0$ exists is clear from (11c). Finally, let s'_i, u'_i , respectively, be the greatest $t < t_i$ and the least $t > t_i$ such that $[s'_i]$ and $[u'_i]$ are on $S(\rho)$, $i = 1, 2$.

Consider the arc $A(\rho)$ of $S(\rho)$ which lies in H . The four points $[s'_i]$, $[u'_i]$ lie on $A(\rho)$ in some linear order. (It is not excluded that certain, or even all, of these points coincide; in which case there will be a corresponding ambiguity in this linear order.) It will be unimportant in which sense this order is established so that the twenty-four permutations on four letters reduce to twelve cases that will be considered distinct. Of these twelve possibilities, four are eliminated, in view of Definition 4, by Assumption A. Then the remaining eight possibilities may be reduced, by making use of the fact that the above notation may be changed, by reversing the direction of the parametrization along C , to one of the following two types:

Case 4.1. The linear order is $[s'_1], \{[s'_2], [u'_2]\}, [u'_1]$.

Case 4.2. The linear order is $\{[s'_1], [u'_1]\}, \{[s'_2], [u'_2]\}$.

Here the curly brackets signify that it is of no consequence which of the two symbols contained occurs first; i. e., $\{p_1, p_2\}$ means either p_1, p_2 or p_2, p_1 .

In Case 4.1 the arc $[s'_1, u'_1]$ separates, in a weak sense, the arc $[s'_2, u'_2]$ from S in $S(\rho)$.¹² In particular, every point of S is as close to some point of $[s'_1, u'_1]$ as to any point of $[s'_2, u'_2]$. Thus there are no points of $P(s'_2, u'_2)$ in S . In view of (11b) this means that $P(s'_2, u'_2)$ is empty; i. e., that (s'_2, u'_2) is non-significant. Then by Lemma 9, (s'_2, u'_2) is linear. But, by (11c) and the fact that $[s'_1, u'_1]$ separates $[s'_2, u'_2]$ from S , this implies that also $[s'_1, u'_1]$ is linear and these two arcs $[s'_i, u'_i]$ are both segments of L . Thus Case 4.1 reduces to Case 1.

Case 4.2 is somewhat more troublesome. Let it be assumed, for the sake of definiteness, that the order is actually $[s'_1], [u'_1], [s'_2], [u'_2]$. It will be clear that this is no restriction since the proof will not refer to the nature of C outside $S(\rho)$.¹³ Now let L be chosen as the Y -axis of a Cartesian (X, Y) plane with origin at the point $[t_1] = [t_2]$. Then either $[s'_1]$ and $[u'_1]$ are both in the closed upper half plane or $[s'_2], [u'_2]$ are both in the closed lower half plane (provided the positive Y direction is chosen appropriately). It will be a notational assumption that the first of these alternatives is true. It will also be assumed that not both $[s'_1]$ and $[u'_1]$ are on the X -axis.¹⁴

Now consider the ellipse with foci at $[s'_1]$ and $[u'_1]$ and passing through $[t_1] = [t_2]$. It is easily seen that this ellipse has, at the point $[t_1] = [t_2]$, a negative slope; i. e., that a point p^* may be chosen on S , in the open second quadrant and in this ellipse. It is supposed that p^* is chosen so that the principal arc $S(p^*, [t_1])$ of S , joining p^* and $[t_1]$ has an angular measure $< \frac{1}{3}\pi$. The fact that p^* is in the given ellipse means that

$$\rho([s'_1], p^*) + \rho(p^*, [u'_1]) < \rho([s'_1], [t_1]) + \rho([t_1], [u'_1]).$$

(cf. Definition 16). Now let l^* be the greatest $t < t_1$ such that $[l^*]$ is on the chord $K(p^*, [s'_1])$ joining p^* and $[s'_1]$. Then it is immediately seen that

$$\rho([l^*], p^*) + \rho(p^*, [u'_1]) < \rho([l^*], [t_1]) + \rho([t_1], [u'_1]).$$

Thus it is seen that the curve C^* which results from C by replacing the arc (l^*, u'_1) by the two chords $K([l^*], p^*) + K(p^*, [u'_1])$ is shorter than C .

It will now be shown that this C^* has the property (ϵ) with respect to M so that C is not an ϵ -ergodic curve for M . This contradiction will complete

¹² In case $[s'_1] = [s'_2], [u'_1] = [u'_2]$ or $[s'_1] = [u'_2], [s'_2] = [u'_1]$, this statement may be understood as a notational assumption, in view of assumption A.

¹³ It is easily seen, by an argument similar to that used in Case 4.1, that the arcs (t_1, u'_1) and (s'_2, t_2) are non-significant and therefore linear, but this fact will not be needed.

¹⁴ This is clearly justified unless all four points s'_i, u'_i coincide on the X -axis. If this occurs one simply chooses a smaller value of ρ in defining $S(\rho)$. If the same trouble occurs for all small $\rho > 0$, then $[t_1] = [t_2]$ is actually in Case 1 which has been treated.

the elimination of Case 4. 2. To this end note first that the arc (t^*, u_1) is separated from S in $S(\rho)$ by the arc $A(s'_1, u'_2)$ defined by

$$A(s'_1, u'_2) = [s'_1, t^*] + K([t^*], p^*) + S(p^*, [t_1]) + [t_2, u'_2]$$

(where $S(p^*, [t_1])$ is the principal arc of S joining p^* and $[t_1]$ as mentioned above). This is true since (t^*, u_1) cannot cross $[s'_1, t^*]$ or $[t_2, u'_2]$ by Assumption A, $K([t^*], p^*)$, by the definition of t^* , or $S(p^*, [t_1])$ by (11c).

Now suppose C^* has not the property (ϵ) with respect to M . Then some point \bar{p} of M at a distance $\leq \epsilon$ from C would be at a distance $> \epsilon$ from C^* . Since C^* contains all of C save (t^*, u_1) , this point \bar{p} , in particular, would have to be salient to (t^*, u_1) . Then, by (11b), $\bar{p} \subset S$. Then the ϵ -circle about the point $\bar{p} \subset S$ has an interior or boundary point on (t^*, u_1) but does not cross C^* and in particular does not cross $A(s'_1, u'_2) - S(p^*, [t_1])$. Thus, according to the preceding paragraph, this ϵ -circle about \bar{p} must cross $S(p^*, [t_1])$ twice. But this is impossible for $\bar{p} \subset S$ since $S(p^*, [t_1])$ has been assumed to have an angular measure $< \frac{1}{2}\pi$. According to the preceding paragraph the treatment of Case 4. 2 (and consequently of Case 4) is complete.

Case 5. There are at least two intersecting salient circles through $[t_1] = [t_2]$. In this case clearly both $[t_1]$ and $[t_2]$ are multisignificant and $\theta(t_1) = \theta(t_2)$. Here two cases are distinguished of which one is trivial.

Case 5. 1: $\theta(t_1) = \theta(t_2) = \pi$. This case reduces immediately to Case 1 in view of Lemma 12 (cf. particularly footnote 9).

Case 5. 2: $\theta(t_1) = \theta(t_2) < \pi$. The treatment of Case 5. 2 closely parallels that of Case 4 and will not be given. It is enough to note that Lemma 4 serves here the purpose served by Lemma 3 in Case 4 while Lemma 12 replaces Lemma 10. Of course, in view of Lemma 4, the four arcs (s_i, t_i) , (t_i, u_i) must always be considered separately, but the only difficulty introduced by this is a notational one.

Case 6. There are exactly two, mutually tangent, salient circles through $[t_1] = [t_2]$. Let p_1, p_2 be the centers of these two salient circles S_1, S_2 and let S^*_1, S^*_2 be the circles of radius $\frac{1}{2}\epsilon$ about p_1, p_2 , respectively. Let s_i, u_i , $i = 1, 2$, be chosen so that

$$(12a) \quad s_1 < t_1 < u_1 < s_2 < t_2 < u_2;$$

$$(12b) \quad \bar{P}(s_1, u_1) + \bar{P}(s_2, u_2) \subset S^*_1 + S^*_2;$$

as is possible by Lemma 3.

Let $S_i(\rho_i)$, $i = 1, 2$, be a circle of center p_i and radius $\epsilon + \rho_i$ where $\rho_i > 0$ is so small that $S_i(\rho_i)$ meets all four arcs (s_1, t_1) , (t_1, u_1) , (s_2, t_2) ,

(t_2, u_2) .¹⁵ Let s_{ij}, u_{ij} , respectively, be the greatest $t < t_i$, and the least $t > t_i$, such that $[s_{ij}]$ and $[u_{ij}]$ are on $S_j(\rho_j)$; $j = 1, 2$; $i = 1, 2$. Finally, let

$$s'_i = \max_{j=1,2} s_{ij}, \quad u'_i = \min_{j=1,2} u_{ij}.$$

Now let the points p_1, p_2 , respectively, be the points $(0, \epsilon)$, $(0, -\epsilon)$ of a Cartesian (X, Y) plane. Suppose that the point $[t_1] = [t_2]$ is not in Case 1 or Case 2. Then there are two cases to be distinguished.

Case 6.1. Two of the arcs (s'_i, t_i) , (t_i, u'_i) lie in the right half plane and two in the left half plane. Since $[t_1] = [t_2]$ is not in Case 1, these four arcs do not all lie on the X -axis. Choose the quadrant which contains a point of one of these arcs as the first quadrant by making the appropriate changes of notation. Let K be a half line terminated by $[t_1] = [t_2]$ and lying in the first quadrant. Then, clearly, K can be chosen so near the positive X -axis that it does not meet S^*_1 or S^*_2 but does meet one of the four arcs (s'_i, t_i) , (t_i, u'_i) . Let t^* be the first point $\neq [t_1]$ where K meets one of these four arcs. (That there is such a first point is clear from the fact that near $[t_1]$, K lies in one of the salient circles S_1, S_2 and so does not meet C .) For the sake of definiteness let it be assumed that $s'_1 < t^* < t_1$ and that $[s'_2, t'_2]$ is the other arc in the right half plane. It is clear from the definition of t^* , together with Assumption A, that $[s'_2, t_2]$ is "below" $[s'_1, t_1]$. Assume also that S^*_1 is in the upper half plane.

Then there are no points of S^*_2 salient to (t^*, t_1) . In fact, the entire arc (s_{12}, t_1) is separated by (s_{22}, t_2) from S^*_2 in the convex curve consisting of that part of $S_2(\rho_2)$ lying below and to the right of that tangent line to S^*_2 through $[t_1] = [t_2]$ which has positive slope. Thus, by (12b), all points $P(t^*, t_0)$ lie in S^*_1 .

Now it may be shown, by an argument exactly like that used in the proof of Lemma 10, that the curve obtained from C by replacing the arc (t^*, t_1) by the chord $K([t^*], [t_1])$ is a shorter curve with the property (ϵ) . This contradiction completes the treatment of Case 6.1.

Case 6.2. At least three of the arcs (s'_i, t_i) , (t_i, u'_i) lie in the same half plane determined by the Y -axis. In this case choose the half plane which contains at least three of these arcs as the right half plane. Since $[t_1] = [t_2]$ is not in Case 2, these three arcs do not lie on the X -axis. Choose the quadrant which contains a point of one of these arcs as the first quadrant. Then all

¹⁵ Again this is possible except in the case that some one of these four arcs lies on the boundary of $S_i(0) = S_i$. The necessary modification in this case is trivial.

the essential features of Case 6.1 are obtained and the treatment proceeds in the same way by simply neglecting the extraneous arc or arcs.

This completes the proof of Theorem 1.

7. Corollaries. Theorem 1 allows several previous results to be strengthened. In connection with the inequality (2c) for $\theta(t)$ we have

LEMMA 13. $\theta(t) = \pi$ if and only if $[t]$ is doubly significant.

Proof. If $\theta(t) = \pi$ for a multisignificant point $[t]$, then, by Lemma 12 (cf. footnote 9), C has a double point. But this is impossible by Theorem 1.

Then Lemma 8 can be stated more simply as

LEMMA 8 bis. *The set of doubly significant points of C is closed.*

8. Local properties resumed. In this section the discussion of the nature of C in the neighborhood of the various types of points will be resumed and further results obtained which are more conveniently proved with the help of Theorem 1. The first of these results is complementary to Lemma 11 and supplementary to Lemmas 10 and 12.

LEMMA 14. *Let $[t_0]$ be not doubly significant. Then some arc (s, u) with $s < t_0 < u$ is convex toward $P(t_0)$.¹⁶*

Proof. If $[t_0]$ is non-significant, this is so by Lemma 9.

Suppose first, then, that $[t_0]$ is simply significant and let s_1, u_1 be chosen so that

$$(13a) \quad s_1 < t_0 < u_1,$$

$$(13b) \quad \bar{P}(s_1, u_1) \subset S(t_0).$$

The possibility of satisfying (13b) follows from Lemma 6. From (13b) it follows, à fortiori, that

$$(14) \quad \bar{P}(t_1, t_2) \subset S(t_0), \quad s_1 \leq t_1 < t_2 \leq u_1.$$

Now let $S(\rho)$ denote a circle of radius $\epsilon + \rho$ about $p(t_0)$, where $\rho > 0$ is so small that $S(\rho)$ cuts both arcs (s_1, t_0) , (t_0, u_1) and let s, u , respectively, be the greatest $t < t_0$ and the least $t > t_0$ such that $[s]$ and $[u]$ are on $S(\rho)$. (The existence of such a $\rho > 0$ is obvious from Lemma 10.) Let $A(s, u)$ be the principal arc of $S(\rho)$ joining $[s]$ and $[u]$.

Then the closed curve Γ defined by

$$\Gamma = [s, u] + A(s, u)$$

¹⁶ An arc is said to be convex if it can be made an arc of a convex curve. A non linear arc is said to be convex toward a given set if the given set necessarily lies outside any such convex curve. A linear segment is said to be convex toward a given set if the set lies on one side of the linear extension of the segment.

is simple. In fact $[s, u]$ cannot touch itself by Theorem 1 and cannot touch $A(s, u)$ (except at $[s]$ and $[u]$) by the definition of s and u . It will now be shown that the simple closed curve Γ is convex. Suppose this is not the case; i. e., that some chord $K([t_1], [t_2])$ lies outside Γ , where $s \leq t_1 < t_2 \leq u$. Then the arc consisting of

$$[s, u] - (t_1, t_2) + K([t_1], [t_2])$$

separates (t_1, t_2) from $S(t_0)$ in $S(\rho)$. But, by a now familiar argument, this contradicts (14). The fact that the convex arc (s, u) is convex toward $p(t_0)$ is obvious from Lemma 10.

The only case remaining to be considered is that $[t_0]$ is multisignificant. In this case (13a) is replaced by a choice of s_1, u_1 such that

$$s_1 < t_0 < u_1, \\ \bar{P}(s_1, t_0) \subset S_1(t_0), \quad \bar{P}(t_0, u_1) \subset S_2(t_0)$$

(where $S_1(t_0), S_2(t_0)$ are, of course, the terminal salient circles through $[t_0]$), by using Lemma 4. Then a double repetition of the preceding argument, somewhat modified of course, shows that some (s, t_0) is convex toward $p_1(t_0)$ and some (t_0, u) is convex toward $p_2(t_0)$. These facts, together with Lemma 12, are easily seen to imply the convexity of (s, u) toward $P(t_0)$.

It should be remarked that this Lemma 14 is not simply a consequence of the results of Lemma 8 bis, Lemma 10, Theorem 1 (as might be suspected) in view of the possible existence of linear segments converging to $[t_0]$. In fact it is quite easy to construct an example of a simple curve with the local half plane or "local supporting line" property of Lemma 10 at every point but which is not locally convex at some point. However, this can only be done by the introduction of linear segments. The complementary Lemmas 14 and 11 lead immediately to

THEOREM 2. *Let M be an arbitrary bounded plane point set. Let $\epsilon > 0$ be fixed and let C be an ϵ -ergodic curve for M . Then at any point $[t]$ of C , $0 < t < 1$, there is a right and left hand tangent to C .*

Proof. If $[t_0]$ is not doubly significant, then, in view of Lemma 14, this follows from a well known theorem on convex curves. If $[t_0]$ is doubly significant, this is trivial in view of Lemma 11.

The difficulty (of linear segments) mentioned above in connection with the extension from the local supporting line to local convexity also prevents the extension from local convexity to convexity in the large. In the case at hand, however, the extension from local convexity to a kind of convexity in the large can be made without eliminating all linear segments—it is enough to eliminate those linear segments which are non-significant.

LEMMA 15. *Let (s, u) be an arc of C containing no non-significant or doubly significant points. Then (s, u) consists of a finite number of convex arcs.*

Proof. Let $\phi(t)$ be the inclination which the left hand tangent to C at $[t]$ has with the x -axis. In view of Lemma 14 the indetermination $(\pm 2k\pi)$ of $\phi(t)$ may be determined so that $\phi(t)$ is monotone (in the weak sense) in some neighborhood of each t for which $[t]$ is not doubly significant. At the same time $\phi(t)$ may be supposed bounded. To see this it is enough to notice that C cannot spirally converge to a point (remembering that C is simple and of finite length). But this is easily seen since if spiral convergence occurred the points of C near the limit point would be non-significant and hence C would be linear near this point which is a contradiction. It is now supposed that some well determined $\phi(t)$ is chosen which is bounded and weakly monotone in some neighborhood of every t where $[t]$ is not doubly significant.

Now let (s, u) contain no non-significant or doubly significant points so that $\phi(t)$ is weakly monotone in some neighborhood of every t for $s < t < u$. It will be shown now that $\phi(t)$ is weakly monotone in the entire interval $s < t < u$. For suppose this is not so. Then there is some sub-interval $s_1 \leq t \leq u_1$ of $s < t < u$ such that the point t_1 where $\phi(t)$ attains its maximum¹⁷ (or its minimum) over the interval $s_1 \leq t \leq u_1$ is an interior point $s_1 < t_1 < u_1$. For the sake of definiteness consider the case of a maximum; i. e., suppose that

$$(15a) \quad s < s_1 < t_1 < u_1 < u;$$

$$(15b) \quad \phi(t_1) \geq \phi(t), \quad s_1 \leq t \leq u_1.$$

But (15b) contradicts the local monotony of $\phi(t)$ unless the equality sign holds in (15b) near t_1 ; at least on one side of t_1 . Thus there are values s_2, u_2 such that¹⁸

$$(16a) \quad s_1 < s_2 \leq t_1 \leq u_2 < u_1, \quad s_2 < u_2;$$

$$(16b) \quad \phi(t_1) = \phi(t), \quad s_2 < t < u_2.$$

Without loss of generality, let it be assumed that the segment $(s_2, u_2) = K([s_2], [u_2])$ lies along the x -axis and that $\phi(t_1) = 0$. By the local monotony of $\phi(t)$ at s_2 and u_2 , in connection with (16b) and footnote 18, values s_3, u_3 may be chosen so that $s_3 < s_2, u_2 < u_3$ and

¹⁷ The fact that $\phi(t)$ actually attains its maximum (minimum) for any such closed interval is a trivial consequence of the local monotony of $\phi(t)$.

¹⁸ It is supposed s_2, u_2 are chosen so that $u_2 - s_2$ is as large as possible. The equation of (16b) may or may not hold with $t = s_2$ or $t = u_2$.

$$(17a) \quad \phi(t) < 0, \quad s_3 < t < s_2;$$

$$(17b) \quad \phi(t) < 0, \quad u_2 < t < u_3.$$

On the other hand, by Lemma 14, these values s_3, u_3 may also be chosen so that

$$(18a) \quad (s_3, s_2) \text{ is convex toward } P(s_2);$$

$$(18b) \quad (u_2, u_3) \text{ is convex toward } P(u_2).$$

Now, by assumption, all points $[t]$ for $s_2 < t < u_2$ are either simply significant or multisignificant. But this last alternative is seen to be impossible if Lemma 12 is compared with (16b). Thus

$$P(t) = (p(t)), \quad s_2 < t < u_2.$$

The set of all points $\{p(t)\}$, for $s_2 < t < u_2$, is, in view of Lemma 6 and Definitions 10 and 12, a linear segment parallel to (s_2, u_2) and lying ϵ units above or ϵ units below (s_2, u_2) . By comparing (17a) and (18a), it is seen, in view of Lemma 6, that this $\{p(t)\}$ lies below (s_2, u_2) . However (17b) and (18b) show, in the same way, that $\{p(t)\}$ lies above (s_2, u_2) . This contradiction completes the proof of the fact that $\phi(t)$ is weakly monotone in the entire interval $s < t < u$.

To complete the proof of Lemma 15, it is enough to divide (s, u) into sub-arcs such that the variation of $\phi(t)$ over any sub-arc is $< \pi$. By the monotony and boundedness of $\phi(t)$ the number of such sub-arcs is finite. That each such sub-arc is convex is then trivial.

As usual a point where the right and left hand tangents differ will be called a corner. In particular, every multisignificant point is a corner. On the other hand a non-significant point or a doubly significant point cannot be a corner. A simply significant point may or may not be a corner as trivial examples show. With regard to corners, in addition to these remarks, there will now be shown:

THEOREM 3. *With the notations of Theorem 2, there are only a countable number of corners on C .*

Proof. Let Q_n, Q_d denote, respectively, the set of non-significant, doubly significant points of C . The set Q_n is, by Lemma 5, an open subset of C so that Q_n is of the form

$$Q_n = \sum_{i=1}^{\infty} (s_i, u_i).$$

But, by Lemma 9, no point of Q_n is a corner. Thus the set

$$\bar{Q}_n = \sum_{i=1}^{\infty} [s_i, u_i]$$

contains at most a countable number of corners.

Now the set Q_d , which is closed by Lemma 8 bis, can contain no corners in view of Lemma 11. Thus to prove this Theorem 3, it is enough to show that there are at most a countable number of corners among the points of

$$Q = C - Q_d - \bar{Q}_n = C - (\bar{Q}_d + \bar{Q}_n).$$

But this Q is an open subset of C ; i. e.,

$$Q = \sum_{i=1}^{\infty} (s'_i, u'_i).$$

Also (s'_i, u'_i) consists of a finite number of convex arcs, by Lemma 15. Thus

$$Q = \sum_{i=1}^{\infty} (s_i'', u_i'')$$

where (s_i'', u_i'') is convex. But it is a standard theorem that a convex arc contains only a countable number of corners. This completes the proof of Theorem 3.

As usual, by a cusp will be meant a point where there is a well defined tangent line but where the curve does not cross the normal. With regard to these it is easy to prove

THEOREM 4. *With the notations of Theorem 2, the curve C has no cusps.*

Proof. In view of Lemma 14, a point $[t]$ which is not doubly significant cannot be a cusp of C , since a convex curve can have no cusps. Thus it is sufficient to show that a doubly significant point cannot be a cusp. This can easily be done by an argument exactly like that which showed that $\theta(t) = \pi$ was impossible for a multisignificant point.

9. The ergodic function. In this section, contrasting with the preceding, C will be used to denote any continuous rectifiable plane Jordan curve of length L while an ϵ -ergodic curve for M will be denoted by $C(\epsilon)$ and its length by $\Lambda(\epsilon)$.

Let $D(C; \epsilon)$ denote the set of all points in the plane at a distance $\leq \epsilon$ from some point of C . Thus $D(C; \epsilon)$ is the domain swept out by a circle of radius ϵ whose center traverses C . Then $D(C; \epsilon)$ is measurable (in fact closed) and

$$\text{meas } D(C; \epsilon) \leq 2\epsilon L + \pi\epsilon^2.$$

This inequality is given by Errera¹⁰ for simple Jordan arcs, but actually, in

¹⁰ *Loc. cit.*, 2.

his proof, no use is made of the fact that the arc is simple. In fact the inequality is strengthened if C is not simple.

The fact that C has the property (ϵ) with respect to M is clearly expressed by

$$(19) \quad D(C; \epsilon) \supset M.$$

But since $D(C; \epsilon)$ is closed, (19) may be strengthened to

$$(20) \quad D(C; \epsilon) \supset \bar{M}$$

where \bar{M} is the closure of M . Then by (19) and (20) and the fact that $L = \Lambda(\epsilon)$ for $C = C(\epsilon)$

$$(21) \quad \text{meas } \bar{M} \leq 2\epsilon\Lambda(\epsilon) + \pi\epsilon^2.$$

Now, immediately, we have

LEMMA 16. *For an arbitrary bounded set M*

$$\liminf_{\epsilon \rightarrow 0} 2\epsilon\Lambda(\epsilon) \geq \text{meas } \bar{M}.$$

Next it is to be shown that $\limsup 2\epsilon\Lambda(\epsilon) \leq \text{meas } \bar{M}$. In this direction we prove first

LEMMA 17. *Let R be a rectangle of perimeter P . Then*

$$2\epsilon\Lambda(\epsilon) \leq \text{meas } R + 2P\epsilon + 16\epsilon^2.$$

Proof. Let the rectangle be placed on a Cartesian coördinate system with its vertices at $(0, 0)$, $(a, 0)$, $(0, b)$, (a, b) . Then let a curve $C = C_\epsilon$ be traced in the following way: First trace the horizontal segment $(0, 0)$ to $(a, 0)$, then the semicircle of radius ϵ which has the segment $(a, 0)$ to $(a, 2\epsilon)$ as diameter and which lies outside R , then the segment $(a, 2\epsilon)$ to $(0, 2\epsilon)$, then the semicircle of radius ϵ which has the segment $(0, 2\epsilon)$ to $(0, 4\epsilon)$ as diameter and which lies outside R , then the segment $(0, 4\epsilon)$ to $(a, 4\epsilon)$, etc. Continue in this manner until a horizontal segment has been drawn which lies above R . It may be easily verified that for the curve C_ϵ so constructed, the inequality (21) becomes an equality; i. e., we have

$$(22) \quad \text{meas } D(C_\epsilon; \epsilon) = 2\epsilon L(\epsilon) + \pi\epsilon^2$$

where $L(\epsilon)$ is the length of C_ϵ . It is equally obvious that

$$(23) \quad R \subset D(C_\epsilon; \epsilon) \subset R^*$$

where R^* is the rectangle with vertices $(-2\epsilon, -2\epsilon)$, $(a + 2\epsilon, -2\epsilon)$, $(-2\epsilon, b + 2\epsilon)$, $(a + 2\epsilon, b + 2\epsilon)$. The first inclusion (23) shows that C_ϵ is ϵ -ergodic to D so that

$$(24) \quad \Lambda(\epsilon) \leq L(\epsilon)$$

and the second inclusion (23) gives

$$(25) \quad \text{meas } D(C_\epsilon; \epsilon) \leq \text{meas } R + 2P\epsilon + 16\epsilon^2.$$

Combination of (22), (24), and (25) gives the inequality of Lemma 17.

We are now ready to demonstrate

LEMMA 18. *For an arbitrary bounded set M*

$$\lim_{\epsilon \rightarrow 0} 2\epsilon \Lambda(\epsilon) \leq \text{meas } \bar{M}.$$

Proof. Let $\eta > 0$ be chosen arbitrarily. Then, since \bar{M} is closed, there exists a finite set of rectangles, R_1, R_2, \dots, R_n , such that

$$(26) \quad \sum_{i=1}^n R_i \supset \bar{M} \supset M$$

and

$$(27) \quad \text{meas } \sum_{i=1}^n R_i \leq \text{meas } \bar{M} + \eta.$$

In each of the rectangles R_i consider an ϵ -ergodic curve $C_i = C_i(\epsilon)$ of length $\Lambda_i(\epsilon)$. Let C_i be oriented so that it is possible to speak of the beginning point of C_i and the end point of C_i . Finally consider the curve $C = C_\epsilon$ consisting of the n curves C_i and the $n-1$ linear segments joining the end point of C_i to the beginning point of C_{i+1} ($i = 1, 2, \dots, n-1$). Then C is clearly ϵ -ergodic to $\sum_{i=1}^n R_i$ and, by (26), à fortiori, to M . Thus

$$(28) \quad \Lambda(\epsilon) \leq L(\epsilon)$$

where $\Lambda(\epsilon)$ is the ergodic function for M and $L(\epsilon)$ is the length of C_ϵ . On the other hand

$$(29) \quad L(\epsilon) \leq \sum_{i=1}^n \Lambda_i(\epsilon) + (n-1)D$$

where D is the maximum distance between two points of $\sum_{i=1}^n R_i$. Applying Lemma 17 to each Λ_i we have from (29)

$$(30) \quad 2\epsilon L(\epsilon) \leq \text{meas } \sum_{i=1}^n R_i + 2\epsilon \sum_{i=1}^n P_i + 16n\epsilon^2 + 2(n-1)\epsilon D$$

where P_i is the perimeter of R_i . Combining (27), (28), and (30) we have at once

$$(31) \quad \limsup_{\epsilon \rightarrow 0} 2\epsilon \Lambda(\epsilon) \leq \text{meas } \bar{M} + \eta.$$

Since (31) holds for an arbitrary $\eta > 0$, the proof of Lemma 18 is complete.

Lemma 16 and Lemma 18 together imply

THEOREM 5. *For an arbitrary bounded set M*

$$\lim_{\epsilon \rightarrow 0} 2\epsilon \Lambda(\epsilon) = \text{meas } \bar{M}.$$

REMARKS ON A SPECIAL CLASS OF ALGEBRAS.*

By O. F. G. SCHILLING.

It was shown by Hasse and Witt that the structure of normal simple algebras over algebraic numberfields and certain fields of algebraic functions can be described in terms of the arithmetic of the underlying groundfield.¹ In this note we discuss algebras over function fields of one variable whose coefficient fields are fields which have only cyclic extensions. It turns out that quite a few of the results of the afore-mentioned theories are still true under our assumptions, e. g. the theorem concerning the sum of the local invariants of an algebra. However, the step from the theory of algebras to class field theory can no more be made. Our results throw some light on the axiomatic treatment of the class field theory in the large. They clearly indicate that the validity of the norm theorem does not imply the law of reciprocity. The reason for this deviation from the classical theory can be found in the fact that the Takagi group of a cyclic extension is in general a proper subgroup of a suitably defined Artin group.

1. Structure of the groundfield. Let T be a field which has only cyclic extensions. We shall suppose that for *every* integer n there exists at least one cyclic extension T_n of degree n over T . The Galois theory then immediately implies that the extensions T_n are unique, i. e. for every integer n there exists exactly one field T_n . We now want to investigate the structure of the field T .

LEMMA 1. *The field T is either an absolutely algebraic field of characteristic $\chi \neq \infty$ whose Steinitz number has no infinite component or it is relatively complete with respect to a non-trivial valuation \tilde{V} .*

Proof. We distinguish two cases

- i) T admits no valuation but the trivial one,
- ii) T has non-trivial valuations V .

* Received August 21, 1939.

¹H. Hasse, "Theorie der relativ-zyklischen algebraischen Funktionenkörper, insbesondere bei endlichem Konstantenkörper," *Journ. f. d. r. u. a. Math.*, vol. 172 (1935), pp. 37-54; E. Witt, "Riemann-Rochscher Satz und Z-Funktion im Hyperkomplexen," *Mathematische Annalen*, vol. 110 (1935), pp. 12-28. These papers will be referred to as H and W, respectively.

In the first case T necessarily is a field of characteristic $\chi \neq \infty$. Moreover, T must be absolutely algebraic over its prime field T_χ for otherwise we could construct non-trivial valuations by means of a transcendence basis.² Let $T_\chi < T^{(1)} < \dots < T^{(i)} < \dots < T = \Sigma T^{(i)}$ be an approximating tower of T over T_χ . Then the formal least common multiple of the degrees $[T^{(i)} : T_\chi]$ is called the Steinitz number of T . The assumption that T have for every integer n exactly one (cyclic) extension T_n implies then that the Steinitz number has no infinite component.³

In the second case the field T admits at least one non-trivial valuation V . If the value group of V is well-ordered in a suitable fashion then it can be shown that V is the composite of a rank 1 valuation \tilde{V} and a valuation V' of the residue class field of \tilde{V} .⁴ Let \tilde{T} be the complete closure of the field T with respect to the valuation \tilde{V} . In order to prove that the field T is relatively complete with respect to the valuation \tilde{V} it suffices to show that \tilde{T} contains no other elements algebraic over T but the elements of T .⁵ In other words, we must prove that T is the universal decomposition field (with respect to \tilde{V}) of its algebraic closure. Let $f_n(x) \equiv x^n + a_1x^{n-1} + \dots + a_n = 0$ be an irreducible equation of degree n with coefficients in \tilde{T} . We associate with $f_n(x)$ a polynomial $g_n(x) = x^n + b_1x^{n-1} + \dots + b_n$ with coefficients in T such that $V(a_i - b_i) > M > 0$, where M is sufficiently large. It can be shown that $g_n(x) = 0$ is irreducible in \tilde{T} and that its roots generate the same field over \tilde{T} as the roots of $f_n(x) = 0$.⁶ Since $g_n(x) = 0$ is also irreducible in T its roots generate the cyclic extension T_n of degree n over T . Hence the field generated by $f_n(x) = 0$ is given as $\tilde{T}T_n$, i. e. it is cyclic and has relative degree n . Thus T is relatively complete with respect to the valuation \tilde{V} . If the valuation \tilde{V} is discrete (i. e., if its value group is isomorphic with the additive group of all integers), then $T = \tilde{T}$ by a theorem of F. K. Schmidt.⁷

In general, we can prove that T is relatively complete with respect to exactly one rank 1 valuation. Namely suppose that T is relatively complete

² A. Ostrowski, "Untersuchungen zur arithmetischen Theorie der Körper," *Mathematische Zeitschrift*, vol. 39 (1934), pp. 269-404.

³ M. Moriya and O. F. G. Schilling, "Zur Klassenkörpertheorie über unendlichen perfekten Körpern," *Journal of the Fac. of Science Hokkaido Imperial University*, Ser. I, vol. 5 (1937), pp. 189-205.

⁴ W. Krull, "Allgemeine Bewertungstheorie," *Journ. f. d. r. u. a. Math.*, vol. 167 (1931), pp. 160-196.

⁵ A. Ostrowski, *loc. cit.*

⁶ O. F. G. Schilling, "A generalization of local class field theory," *American Journal of Mathematics*, vol. 60 (1938), pp. 667-704.

⁷ F. K. Schmidt, "Mehrfach perfekte Körper," *Mathematische Annalen*, vol. 108 (1933), pp. 1-25.

with respect to another rank 1 valuation V_1 . Then we can construct irreducible equations $h_n(x) = 0$ with coefficients in T which have prescribed characters of decomposition with respect to \tilde{V} and V_1 .⁸ Repeating the preceding argument we immediately see that the existence of a valuation V_1 with the asserted properties leads to a contradiction to the assumptions on the field T .

Remark. For the actual construction of fields T see a paper of the author on formal power series of several variables.⁹

DEFINITION. A field T is said to be quasi-algebraically closed if it is never the center of proper division algebras of finite rank.¹⁰

THEOREM 1. A field T which has only cyclic extensions is quasi-algebraically closed.

Proof. First, our hypothesis implies that the field T is algebraically perfect. Namely T is supposed to possess only cyclic extensions. Now it follows immediately, by a theorem of Albert, that T never is the center of a proper division algebra of degree χ^m , $\chi \neq \infty$.¹¹ Thus it remains to discuss algebras A whose degrees $n = \prod_{i=1}^r p_i^{a_i}$ are relatively prime to the characteristic.

The structure theory of algebras yields that $A \sim D_1 \times \cdots \times D_r$ where the algebras D_i are normal division algebras of degrees $p_i^{b_i}$ over T , respectively; $0 < b_i \leq a_i$. We want to show $D_i \sim T$. Let $p = p_i$. Since T has, by hypothesis, only cyclic extensions, it follows that D_i is split by a field $T^{(j)}$ of degree p^j , $j \leq b_i$. Let $T < T^{(1)} < \cdots < T^{(i+1)} < T^{(i)} < \cdots < T^{(j)}$ be the chain of cyclic subfields of $T^{(j)}/T$ such that $[T^{(i)}:T^{(i-1)}] = p$. We shall prove by induction that $D_i \times T^{(j)} \sim T^{(j)}$ implies $D_i \sim T$. Suppose that we already proved $D_i \times T^{(i)} \sim T^{(i)}$. Then $D_i \times T^{(i-1)} \sim (T^{(i)}/T^{(i-1)}, a_{i-1})$, $a_{i-1} \neq 0$ in $T^{(i-1)}$. If a_{i-1} is a p -th power nothing has to be proved. So let $a_{i-1} \neq c_{i-1}^p$. Suppose that $T^{(i-1)}$ contains the p -th roots of unity. Then $T^{(i)} = T^{(i-1)}(b_{i-1}^{1/p})$. Consequently,

$$(T^{(i)}/T^{(i-1)}, a_{i-1}) \sim (T^{(i-1)}(a_{i-1}^{1/p})/T^{(i-1)}, b_{i-1})^{-1}.$$

⁸ B. L. van der Waerden, *Moderne Algebra*, vol. I (Berlin, 1937), 2nd edition, pp. 201-202.

⁹ O. F. G. Schilling, "Arithmetic in fields of formal power series in several variables," *Annals of Mathematics*, vol. 38 (1937), pp. 551-576.

¹⁰ O. F. G. Schilling, "The structure of local class field theory," *American Journal of Mathematics*, vol. 60 (1938), pp. 75-100.

¹¹ A. A. Albert, "Normal division algebras of degree p^e over fields of characteristic p ," *Transactions of the American Mathematical Society*, vol. 39 (1936), pp. 183-188.

This similarity implies that $a_{I-1} = b_{I-1}d^{p_{I-1}}$, i. e. $D_i \times T^{(I-1)} \sim T^{(I-1)}$. The preceding argument can also be applied for $p = 2$ for our assumptions exclude that T is a totally real field. Suppose next that $T^{(I-1)}$ does not contain the p -th roots of unity ξ^λ , $\lambda = 1, \dots, p$. Consider then the algebra $D_i \times T^{(I-1)} \times T^{(I-1)}(\xi)$ over $T^{(I-1)}(\xi)$ as groundfield. Since $[T^{(I-1)}(\xi) : T] = p - 1$, it follows that $T^{(I)}(\xi)$ is a splitting field of the extended algebra. As before we conclude that $T^{(I-1)}(\xi)$ is a splitting field of $D_i \times T^{(I-1)}(\xi)$. But this is impossible if $D_i \times T^{(I-1)} \not\sim T^{(I-1)}$.

2. Foundations of local class field theory of discrete complete fields.

Let C be a field which is complete with respect to a rank 1 valuation \mathfrak{p} and has the field T as residue class field. Since Hensel's Lemma holds for C it follows that the unramified extensions C_n of degree n over C are in $(1 - 1)$ -correspondence with the extensions T_n of T . Consequently the generating automorphisms F_n of the various Galois groups $G(C_n|C)$ can be selected such that they induce the generating automorphisms of the Galois groups $G(T_n|T)$. Let C_∞ denote the maximal unramified extension of C . The Galois group $G(C_\infty|C)$ is an ideal cyclic group. Selecting once and for all an element F in $G(C_\infty|C)$ we observe that the infinite cyclic group $\{F^\lambda, \lambda \text{ running over the additive group of all integers}\}$ is everywhere dense in $G(C_\infty|C)$. Consequently the element F induces for every n a generating automorphism F_n of $G(C_n|C)$.¹² Having fixed the automorphism F , the substitutions F_n have the same algebraic properties as the Frobenius automorphisms of the classical ramification theory.

In order to derive the local class field theory relative to the field C it is sufficient to prove the following lemma.

LEMMA 2. *All units of C are norms of units in C_n .*

Proof. Let u be an arbitrary unit of C . Then, by Theorem 1, its residue class $u \bmod \mathfrak{p}$ in T is the norm of an element R of T , i. e. $u \equiv NR \pmod{\mathfrak{p}}$. Thus we have a first \mathfrak{p} -adic approximation of u as a norm of a unit $U \equiv R \pmod{\mathfrak{p}}$ in C . Since $u(NU)^{-1} \equiv 1 \pmod{\mathfrak{p}}$, the customary procedure of \mathfrak{p} -adic approximation yields that $u(NU)^{-1} = NH$, where $H \equiv 1 \pmod{\mathfrak{p}}$.¹³

The usual arguments of local class field theory imply that Lemma 2 yields the following theorem.

THEOREM 2. *Every normal simple algebra A over C is similar to a*

¹² O. F. G. Schilling, "Regular normal extensions over complete fields," To appear in the *Annals of Mathematics*.

¹³ E. Witt, "Schiefkörper über diskret bewerteten Körpern," *Journ. f. d. r. u. a. Math.*, vol. 176 (1937), pp. 153-156.

cyclic algebra $(C_n/C, F_n, \pi)^v$ where π denotes a fixed prime element of the valuation \mathfrak{p} .¹⁴

As in the classical theory we define the residue $v/n \pmod 1$ as the invariant of the algebra A . Having selected F and π every algebra A is uniquely determined in its class by its invariant.

3. Algebras in the large. Let k be an arbitrary function field of one variable with coefficients in the field T . Now let A be an arbitrary normal simple algebra over k as groundfield. Since k is a function field of one variable it follows, by a theorem of Tsen, that the algebraically closed field T_∞ of T suffices as a splitting field of A when adjoined to k .¹⁵ Hence a suitable finite extension kT_n of k already splits the given algebra A ,

$$A \times T_n k \sim T_n k.$$

Thus, $A \sim (T_n k/k, F_n, a)$, $a \neq 0$ in k .

We next want to determine the local invariants $r(\mathfrak{p})$ of the algebra A/k .¹⁶ These characters $r(\mathfrak{p})$ of A are uniquely determined by virtue of Theorem 2. First let us observe that

$$(kT_n/k, F_n, a) \sim (kT_n/k, F_n, ab) \text{ for any } b \neq 0 \text{ in } T.$$

Namely, $(kT_n/k, F_n, b) \sim (T_n/T, F_n, b) \times k \sim k$. Consequently the structure of A depends only on the divisor $(a) = \prod_{i=1}^s \mathfrak{p}_i^{a_i}$. Since $A \sim (kT_n/k, F_n, a)$, we get

$$A_{\mathfrak{p}} \sim (T_n k_{\mathfrak{p}}/k_{\mathfrak{p}}, F_n^d, a^e)$$

where $d = (n, f(\mathfrak{p}))$ and $e = f(\mathfrak{p})d^{-1}$. Here $f(\mathfrak{p})$ denotes the absolute degree of the prime divisor \mathfrak{p} . As a consequence of Lemma 2 and Theorem 2 we find that the algebra $A_{\mathfrak{p}}$ is completely determined by the invariant $r(\mathfrak{p}) \equiv f(\mathfrak{p})\alpha(\mathfrak{p})n^{-1} \pmod 1$, where $\mathfrak{p}^{a(\mathfrak{p})} \nmid (a)$. We remark that $r(\mathfrak{p}) \equiv 0 \pmod 1$ if $\mathfrak{p} \nmid (a)$; namely then a is a unit for the prime divisor \mathfrak{p} . Hence, by Lemma 2, $A_{\mathfrak{p}} \sim k_{\mathfrak{p}}$. Therefore, the algebra A is ramified at most at the prime divisors of (a) . As usual it follows that

$$\sum_{(\mathfrak{p})} r(\mathfrak{p}) \equiv 0 \pmod 1 \text{ for the invariants of an arbitrary algebra } A/k,$$

¹⁴ C. Chevalley, "La théorie du symbole de restes normiques," *Journ. f. d. r. u. a. Math.*, vol. 169 (1933), pp. 140-157.

¹⁵ Ch. C. Tsen, "Divisionsalgebren über Funktionenkörpern," *Nach. v. d. Gesell. d. Wiss. Göttingen* (1933), pp. 335-339.

¹⁶ The local invariants $r(\mathfrak{p})$ are defined to be the invariants of the limit algebras $A_{\mathfrak{p}}$ of A .

for $\sum_{i=1}^s f_i(p) \alpha_i(p) = 0$.¹⁷ Thus, we established a generalization of the classical norm theorem.

Suppose now that k is a rational function field $T(x)$. Then the prime divisors p at finite distance with respect to x can be represented by irreducible polynomials π of degree $f(p)$ with respect to x . Moreover, every divisor of degree 0 in $k = T(x)$ is a principal divisor, i. e. it belongs to an element of k . We then can prove that for every finite set p_1, \dots, p_s of prime divisors to which there are associated rational fractions $a_i b_i^{-1}$ whose sum is 0, there exists an algebra $(T_n k/k, F_n, a)$ whose local invariants $r(p_i) = a_i b_i^{-1}$, $r(p) = 0$ if $p \neq p_1, \dots, p_s$.

To prove this assertion we proceed as follows.¹⁸ Put $n = \prod_{i=1}^s b_i f(p_i)$ and $\alpha_i = a_i n (b_i f(p_i))^{-1}$, $i = 1, 2, \dots, s$. Then the divisor $\prod_{i=1}^s p_i^{\alpha_i}$ has the order

$$\sum_{i=1}^s \alpha_i f(p_i) = \sum_{i=1}^s a_i n (b_i f(p_i))^{-1} f(p_i) = n \sum_{i=1}^s a_i b_i^{-1} = 0.$$

Consequently, by assumption, $\prod_{i=1}^s p_i^{\alpha_i} = (a)$. Hence the algebra $(T_n k/k, F_n, a)$ obviously has all the required properties.

If the genus of k is greater than 0 then one can readily construct examples for which there exist no algebras with prescribed invariants. Namely, take for k a field of genus > 0 whose defining equation $f(x, y) = 0$ has coefficients in the field of all complex numbers C . There exist then infinitely many divisor classes whose orders are infinite. Selecting the p_i and $a_i b_i^{-1}$ appropriately one easily can construct the necessary counter examples.

We now want to prove that every division algebra

$$D \sim (kT_n/k, F_n, a) \text{ is ramified.}$$

Let $a = \prod_{i=1}^s \pi_i^{\alpha_i}$, $0 < \alpha_i < n$, where the π_i are irreducible polynomials in x belonging as uniformizing variables to the (finite) prime divisors p_i . Then the algebra $(kT_n/k, F_n, a)$ is similar to the direct product of the s algebras

$$A_i = (kT_n/k, F_n, \pi_i^{\alpha_i}).$$

Each one of these algebras is at most ramified at p_i and p_∞ , where p_∞ denotes the denominator of x . The invariants $r(p_i)$ of A are the same as the invariants of A_i at p_i according to the structure of local algebras. Thus, a finite prime divisor p_i gives rise to a local division algebra, if and only if

¹⁷ H, p. 45.

¹⁸ W, Theorem 18, p. 27.

$$\alpha_i f(p_i) \not\equiv 0 \pmod{n}.$$

Now let us prove that $\alpha_i f(p_i) \equiv 0 \pmod{n}$ implies $A_i \sim k$.

Let $(T_n k/k, F_n, \pi^a)$ be such an algebra. Denote (f, n) by d . Then

$$p = \mathfrak{P}_1 \cdots \mathfrak{P}_d$$

with prime divisors \mathfrak{P}_i in $T_n k$. All these prime divisors $\mathfrak{P}_i = (\Pi_i)$ are principal for $T_n k = T_n(x)$. We have

$$N\mathfrak{P}_i = p^{n/d} = (\pi)^{n/d}.$$

Next $N\mathfrak{P}_i^a = (\pi)^{na/d}$. Hence $(\pi)^a = N(\Pi_i)^\beta$. Namely, there exist integers μ, ν , such that $nd^{-1}\mu + fd^{-1}\nu = 1$. Now, as a consequence of $\alpha f \equiv 0 \pmod{n}$, $\alpha f = gn$. Whence we get $\alpha fd^{-1} = gnd^{-1}$. Consequently,

$$N(\Pi_i)^\nu = (\pi)^{gn/d} = (\pi)^{\alpha f/d}, \text{ and}$$

$$N(\Pi_i)^\mu = (\pi)^{an/d}.$$

Hence

$$N(\Pi_i)^{\nu\mu} N(\pi_i)^{\mu a} = (p)^a = N(\Pi_i)^\beta.$$

Since units are irrelevant for the structure of factor sets in cyclic algebras, we get

$$\pi^a = N\Pi_i^\beta, \text{ or}$$

$$(T_n k/k, F_n, \pi^a) \sim k \text{ if } \alpha f \equiv 0 \pmod{n}.$$

Consequently, the algebras A_i for which $\alpha_i f(p_i) \equiv 0 \pmod{n}$ can be omitted in the representation of the algebra A . Combining these results we find that a cyclic product A which is similar to a proper division algebra over k must have at least two ramifications.

As usual we have ¹⁹

THEOREM 3. *The class group of normal algebras over $k = T(x)$ is isomorphic with a subgroup S of the additive group $\{r(p)\}$ of all vectors of rational numbers mod 1. The group S consists of all vectors for which $\sum_{(p)} r(p) \equiv 0 \pmod{1}$ and $r(p) = 0$ for almost all p .*

Finally, we remark that the index and exponent of any normal algebra over $T(x)$ coincide.²⁰

UNIVERSITY OF CHICAGO.

¹⁹ W, Theorem 19, p. 27.

²⁰ W, Theorem 20, p. 28.

ON A CERTAIN PARTITION FUNCTION.*

By IVAN NIVEN.**

1. Introduction. It has been shown by Schur¹ (and by Gleissberg² with a different method) that the number a_m of partitions of an integer m with summands of the form $6n \pm 1$ equals the number of partitions of m such that the difference between any two summands is at least three, and at least six in case both summands are divisible by three. The purpose of the present paper is the evaluation of the a_m which may be considered as the coefficients of the powers of x in the expansion of the function

$$(1.1) \quad F(x) = \frac{f(x)f(x^6)}{f(x^2)f(x^3)} = \sum_{m=0}^{\infty} a_m x^m$$

as a power series, where

$$(1.2) \quad f(x) = \prod_{j=1}^{\infty} (1 - x^j)^{-1}.$$

That a_m represents the number of partitions of the integer m having summands of the form $6n \pm 1$ is immediately verified by expanding $F(x)$.

The method used is essentially that employed by Professor Rademacher³ in his investigation of the modular function $J(\tau)$. The author takes this opportunity to thank Professor Rademacher for suggesting the problem and for advice on its solution.

2. Transformation formulas. We employ the familiar transformation formula⁴

$$(2.1) \quad f \left\{ \exp \left(2\pi i \frac{iz + h}{k} \right) \right\} \\ = \omega_{h,k} \sqrt{z} \exp \left\{ \frac{\pi}{12k} \left(\frac{1}{z} - z \right) \right\} f \left\{ \exp \left(2\pi i \frac{(i/z) + h'}{k} \right) \right\},$$

* Received May 5, 1939.

** Harrison Research Fellow.

¹ I. Schur, "Zur additiven Zahlentheorie," *Sitzungsberichte der Berliner Akademie*, 1926, pp. 488-495.

² "Über einen Satz von Herrn I. Schur," *Mathematische Zeitschrift*, vol. 28 (1928), pp. 372-382.

³ Hans Rademacher, "The Fourier coefficients of the modular invariant $J(\tau)$," *American Journal of Mathematics*, vol. 60 (1938), pp. 501-512.

⁴ G. H. Hardy and S. Ramanujan, "Asymptotic formulae in combinatory analysis," *Proceedings of the London Mathematical Society* (2), vol. 17 (1918), pp. 75-115, Lemma 4.31.

in which

$$(2.2) \quad (h, k) = 1, \quad hh' \equiv -1 \pmod{k},$$

and $\omega_{h,k}$ is a root of unity frequently used in modular function theory. Hardy and Ramanujan⁵ give the values

$$(2.3) \quad \omega_{h,k} = (-k|h) \exp(-\pi i \{ \frac{1}{4}(2 - hk - h) + \frac{1}{12}(k - 1/k)(2h - h' + h^2h') \})$$

for h odd, and

$$(2.4) \quad \omega_{h,k} = (-h|k) \exp(-\pi i \{ \frac{1}{4}(k - 1) + \frac{1}{12}(k - 1/k)(2h - h' + h^2h') \})$$

for k odd. We wish to obtain a transformation formula similar to (2.1) for the function

$$(2.5) \quad F \left\{ \exp \left(2\pi i \frac{iz + h}{k} \right) \right\},$$

by applying (2.1) to (1.1). There arise four cases, according as $(k, 6)$ has the value 6, 3, 2, or 1, and the value of the function (2.5) is, respectively,

$$(2.6) \quad \Omega_{h,k} \psi_k(z) F \left\{ \exp \left(2\pi i \frac{(i/z) + h'}{k} \right) \right\},$$

$$(2.7) \quad \Omega_{h,k} \psi_k(z) F^{-1} \left\{ \exp \left(\pi i \frac{(i/z) + h'}{k} \right) \right\},$$

$$(2.8) \quad \Omega_{h,k} \psi_k(z) F^{-1} \left\{ \exp \left(2\pi i \frac{(i/z) + h'}{3k} \right) \right\}, \text{ and}$$

$$(2.9) \quad \Omega_{h,k} \psi_k(z) F \left\{ \exp \left(\pi i \frac{(i/z) + h'}{3k} \right) \right\}.$$

In the last three cases, h' is a solution of the congruence $hh' \equiv -1 \pmod{k}$ such that it is divisible by 2, 3, and 6 respectively; clearly this is possible because of the divisibility properties of k in the various cases. Also we have

$$(2.10) \quad \Omega_{h,k} = \frac{\omega_{h,k} \omega_{h,k/6}}{\omega_{h,k/3} \omega_{h,k/2}}, \quad \psi_k(z) = \exp \left\{ \frac{\pi}{6k} (1/z - z) \right\} \text{ for } (k, 6) = 6,$$

$$(2.11) \quad \Omega_{h,k} = \frac{\omega_{h,k} \omega_{2h,k/3}}{\omega_{h,k/3} \omega_{2h,k}}, \quad \psi_k(z) = \exp \left\{ -\frac{\pi}{12k} (1/z + 2z) \right\} \text{ for } (k, 6) = 3,$$

$$(2.12) \quad \Omega_{h,k} = \frac{\omega_{h,k} \omega_{3h,k/2}}{\omega_{h,k/2} \omega_{3h,k}}, \quad \psi_k(z) = \exp \left\{ -\frac{\pi}{6k} (1/3z + z) \right\} \text{ for } (k, 6) = 2, \text{ and}$$

$$(2.13) \quad \Omega_{h,k} = \frac{\omega_{h,k} \omega_{6h,k}}{\omega_{2h,k} \omega_{3h,k}}, \quad \psi_k(z) = \exp \left\{ -\frac{\pi}{12k} (-(1/3z) + 2z) \right\} \text{ for } (k, 6) = 1.$$

⁵ *Loc. cit.*, p. 85.

3. Applying Cauchy's Integral Formula to (1.1) we obtain

$$(3.1) \quad a_m = \frac{1}{2\pi i} \int_C \frac{F(x)}{x^{m+1}} dx = \sum'_{\substack{h,k \\ 0 \leq h < k \leq N}} \frac{1}{2\pi i} \int_{\xi_{h,k}} \frac{F(x)}{x^{m+1}} dx$$

wherein Σ' means that h is summed over values prime to k . We choose C to be the circle $|x| = \exp(-2\pi N^{-2})$, so that the Farey arc $\xi_{h,k}$ is given by

$$x = \exp\left(-2\pi N^{-2} + \frac{2\pi i h}{k} + 2\pi i \phi\right),$$

where $-\theta'_{h,k} \leq \phi \leq \theta''_{h,k}$; and, if $h_1/k_1, h_2/k_2$ are the neighbours on the left and right respectively of h/k in the Farey series of order N ,

$$(3.2) \quad \theta'_{h,k} = -\frac{1}{k(k_1 + k)}, \quad \theta''_{h,k} = \frac{1}{k(k_2 + k)}.$$

Thus we have

$$a_m = \sum'_{\substack{h,k \\ 0 \leq h < k \leq N}} \int_{-\theta'}^{\theta''} F \left\{ \exp\left(\frac{2\pi i h}{k} - 2\pi(N^{-2} - i\phi)\right) \right\} \\ \times \exp\left\{-m\left(-2\pi N^{-2} + \frac{2\pi i h}{k} + 2\pi i \phi\right)\right\} d\phi,$$

the subscripts being omitted from $\theta'_{h,k}$ and $\theta''_{h,k}$. Now set $w = N^{-2} - i\phi$ and $z = kw$.

$$(3.3) \quad a_m = \sum'_{\substack{h,k \\ 0 \leq h < k \leq N}} \exp\left(-\frac{2\pi i h m}{k}\right) \\ \times \int_{-\theta'}^{\theta''} F \left\{ \exp\left(\frac{2\pi i h}{k} - \frac{2\pi z}{k}\right) \right\} \exp(2\pi m w) d\phi.$$

In order to make use of the formulas (2.6), (2.7), (2.8), and (2.9), we break a_m above into the parts $a_m^{(6)}$, $a_m^{(3)}$, $a_m^{(2)}$, and $a_m^{(1)}$ according as $(k, 6)$ equals 6, 3, 2, 1 respectively. Applying (2.6) to $a_m^{(6)}$, we have

$$(3.4) \quad a_m^{(6)} = \sum'_{\substack{h,k \\ 0 \leq h < k \leq N \\ (k,6)=6}} \exp\left(-\frac{2\pi i h m}{k}\right) \Omega_{h,k} \\ \times \int_{-\theta'}^{\theta''} \psi_k(kw) \sum_{n=0}^{\infty} a_n \exp\left(\frac{2\pi i h' n}{k} - \frac{2\pi n}{k^2 w} + 2\pi m w\right) d\phi.$$

Splitting off the first term in the summation in the integrand, we write, making use of the fact that $a_0 = 1$, and (2.10),

$$(3.5) \quad I_1 = \sum_{\substack{k=1 \\ (k,6)=6}}^N \sum'_{h \bmod k} \exp\left(-\frac{2\pi i h m}{k}\right) \Omega_{h,k} \\ \times \int_{-\theta'}^{\theta''} \exp\left(\frac{\pi}{6k^2 w} + 2\pi m w - \frac{\pi w}{6}\right) d\phi.$$

Hence $a_m^{(6)} = I_1 + I_2$, I_2 being the same as (3.4) with the summation in the integrand ranging from 1 to ∞ .

4. The evaluation of I_2 . Formulas (2.3) and (2.10) imply

$$(4.1) \quad \Omega_{h,k} = \exp \left\{ \frac{\pi i}{12} \left(\frac{2h}{k} - \frac{2h'}{k} + \frac{2hk}{3} + \frac{h'k}{3} \right) \right\} \text{ for } (k, 6) = 6,$$

where h' has been chosen so that $hh' \equiv -1 \pmod{12k}$. Using

$$(4.2) \quad -\theta' = -\frac{1}{k(k_1+k)} \leq -\frac{1}{k(N+k)} < \frac{1}{k(N+k)} \leq \frac{1}{k(k_2+k)} = \theta''$$

we may write

$$(4.3) \quad I_2 = \sum_{\substack{k=1 \\ (k,6)=6}}^N \sum_{n=1}^{\infty} a_n \int_{-\frac{1}{k(N+k)}}^{\frac{1}{k(N+k)}} \exp \left\{ -\frac{\pi}{k^2 w} (2n - \frac{1}{6}) + \pi w (2m - \frac{1}{6}) \right\} d\phi \\ \cdot \sum'_{h \bmod k} \exp \left\{ -\frac{2\pi i}{k} (mh - nh') \right\} \Omega_{h,k} \\ + \sum_{\substack{k=1 \\ (k,6)=6}}^N \sum_{n=1}^{\infty} a_n \sum'_{h \bmod k} \exp \left\{ -\frac{2\pi i}{k} (mh - nh') \right\} \Omega_{h,k} \sum_{l=k_1+k}^{N+k-1} \int_{-\frac{1}{kl}}^{\frac{1}{k(l+1)}} \\ + \sum_{\substack{k=1 \\ (k,6)=6}}^N \sum_{n=1}^{\infty} a_n \sum'_{h \bmod k} \exp \left\{ -\frac{2\pi i}{k} (mh - nh') \right\} \Omega_{h,k} \sum_{l=k_2+k}^{N+k-1} \int_{\frac{1}{kl}}^{\frac{1}{k(l+1)}} \\ = S_1 + S_2 + S_3,$$

where the integrand is the same in all three expressions. By (4.1) the inner sum in S_1 is the Kloosterman sum

$$(4.4) \quad \sum'_{h \bmod k} \exp \left\{ -\frac{2\pi i}{12k} \left\{ h \left(12m - 1 - \frac{k^2}{3} \right) + h' \left(-12n + 1 - \frac{k^2}{6} \right) \right\} \right\}.$$

The quantities in parentheses are integers since k is divisible by 6. Since we required $hh' \equiv -1 \pmod{12k}$, this sum may be considered as an incomplete sum mod $12k$; using a device of Estermann,⁶ and an estimate of Salié,⁷ the sum (4.4) is

$$O(k^{2/3+\epsilon}(12m-1-k^2/3, k)^{1/3}) = O(k^{2/3+\epsilon}m^{1/3}).$$

⁶ T. Estermann, "Vereinfachter Beweis eines Satzes von Kloosterman," *Abhandlungen Hamburg. Math. Seminar*, vol. 7 (1929), p. 94.

⁷ H. Salié, "Zur Abschätzung der Fourierkoeffizienten ganzer Modulformen," *Mathematische Zeitschrift*, vol. 36 (1933), p. 264.

Using the inequality

$$(4.5) \quad R \left\{ \frac{\pi}{k^2 w} (2n - \frac{1}{6}) \right\} > R \left\{ \frac{\pi n}{k^2 w} \right\} \\ = \frac{\pi n N^{-2}}{k^2 (N^{-4} + \phi^2)} \geq \frac{\pi n}{k^2 N^{-2} + N^2 k^2 \theta_{h,k}^2} \geq \frac{\pi n}{2}$$

we obtain

$$S_1 = O \left(\sum_{k=1}^N \sum_{n=1}^{\infty} a_n \cdot \frac{2}{kN} \exp \left\{ -\frac{\pi n}{2} + \pi N^{-2} (2m - \frac{1}{6}) \right\} k^{2/3+\epsilon} m^{1/3} \right) \\ = O \left(\frac{1}{N} \exp(2\pi m N^{-2}) m^{1/3} \sum_{n=1}^{\infty} a_n \exp \left(-\frac{\pi n}{2} \right) \sum_{k=1}^N k^{-1/3+\epsilon} \right) \\ = O \left(\frac{1}{N} \exp(2\pi m N^{-2}) m^{1/3} N^{2/3+\epsilon} \right)$$

$$(4.6) \quad S_1 = O(\exp(2\pi m N^{-2}) m^{1/3} N^{-1/3+\epsilon}).$$

Because of the similarity of S_2 and S_3 , we shall treat only the latter. Interchanging the summations with respect to h and l we get

$$S_3 = \sum_{\substack{k=1 \\ (k,6)=6}}^N \sum_{n=1}^{\infty} a_n \sum_{l=N+1}^{N+k-1} \int_{\frac{1}{k(l+1)}}^{\frac{1}{kl}} \exp \left\{ -\frac{\pi}{k^2 w} (2n - \frac{1}{6}) + \pi w (2n - \frac{1}{6}) \right\} d\phi \\ \cdot \sum'_{\substack{h \bmod k \\ N < k+k_2 \leq l}} \exp \left\{ -\frac{2\pi i}{12k} \right\} h \left(12m - 1 - \frac{k^2}{3} \right) + h' \left(-12n + 1 - \frac{k^2}{6} \right) \Big\} \Big\}.$$

In order to interpret the restriction on k_2 in the inner sum, we recall from the theory of Farey series that if

$$\frac{h_1}{k_1} < \frac{h}{k} < \frac{h_2}{k_2}$$

are three neighbors in the Farey series of order N , then

$$hk_1 - h_1k = 1 = h_2k - hk_2$$

so that

$$hk_1 \equiv -hk_2 \equiv 1 \pmod{k},$$

or, by applying (2.2)

$$(4.7) \quad -k_1 \equiv k_2 \equiv h' \pmod{k}.$$

From this it follows that the above restriction on k_2 implies a restriction on h' to an interval mod k equivalent to one or two intervals in the range $0 \leq h' < k$. If the Kloosterman sum is considered mod $12k$, we have a restriction on both h and h' . We proceed to remove the restriction on h , so that the sum may be treated as was (4.4).

We shall prove that if the sum

$$(4.8) \quad \sum_{\substack{h \bmod k \\ N < k+k_2 \leq l}} \exp \left\{ -\frac{2\pi i}{12k} \left\{ h \left(12m - 1 - \frac{k^2}{3} \right) + h' \left(-12n + 1 - \frac{k^2}{6} \right) \right\} \right\}$$

is multiplied by

$$(4.9) \quad 1 + \sum_{a=1}^{11} \exp \left\{ -\frac{2\pi i}{12k} \left\{ \alpha k \left(12m - 1 - \frac{k^2}{3} \right) + \beta k \left(-12n + 1 - \frac{k^2}{6} \right) \right\} \right\}$$

where $\beta = \alpha$ or 7α according as $(k, 12) = 12$ or 6 , the product is

$$(4.10) \quad \sum_{\substack{h \bmod 12k \\ N < k+k_2 \leq l}} \exp \left\{ -\frac{2\pi i}{12k} \left\{ h \left(12m - 1 - \frac{k^2}{3} \right) + h' \left(-12n + 1 - \frac{k^2}{6} \right) \right\} \right\}$$

where h' satisfies the congruence $hh' \equiv -1 \pmod{12k}$. We must prove that

$$(h + \alpha k)(h' + \beta k) \equiv -1 \pmod{12k},$$

or, multiplying by h ,

$$\beta h^2 - \alpha + \alpha \beta h k \equiv 0 \pmod{12}.$$

If $(k, 12) = 12$, we use $\beta = \alpha$ and require

$$\alpha(h^2 - 1) \equiv 0 \pmod{12}$$

which is obviously true since h is prime to k which is divisible by 6 . If $(k, 12) = 6$, we use $\beta = 7\alpha$ and require

$$\alpha(7h^2 - 1) + 7\alpha^2 h k \equiv 0 \pmod{12}$$

which reduces, since $7h^2 - 1 \equiv 6 \pmod{12}$, to

$$\alpha(6 + 7\alpha k h) \equiv 0 \pmod{12}.$$

This is true since the factor in parentheses is divisible by 6 when α is even, and by 12 when α is odd. The argument following (4.4) shows that (4.10), and hence (4.8), equals

$$(4.11) \quad O((12k)^{2/3+\epsilon} (12m - 1 - k^2/3, 12k)^{1/3}) = O(k^{2/3+\epsilon} m^{1/3}).$$

Thus

$$\begin{aligned} S_3 &= O \left(\sum_{k=1}^N \sum_{n=1}^{\infty} a_n \sum_{l=N+1}^{N+k-1} \left(\frac{1}{kl} - \frac{1}{k(l+1)} \right) \right. \\ &\quad \times \exp \left\{ -\frac{\pi n}{2} + \pi N^{-2} (2m - 1/6) \right\} k^{2/3+\epsilon} m^{1/3} \Big) \\ &= O \left(\sum_{k=1}^N \frac{1}{kN} \exp(2\pi m N^{-2}) k^{2/3+\epsilon} m^{1/3} \right), \end{aligned}$$

$$(4.12) \quad S_3 = O(\exp(2\pi m N^{-2}) m^{1/3} N^{-(1/3)+\epsilon}).$$

Combining (4.3), (4.6), and (4.12), we obtain

$$(4.13) \quad I_2 = O(\exp(2\pi m N^{-2}) m^{1/3} N^{-(1/3)+\epsilon}).$$

5. The evaluation of I_1 . We divide the expression I_1 in (3.5) into three parts $Q_0(m)$, $Q_1(m)$, and $Q_2(m)$ by splitting the limits of integration into the parts

$$\left(-\frac{1}{k(N+k)}, \frac{1}{k(N+k)}\right), \left(-\frac{1}{k(k_1+k)}, -\frac{1}{k(N+k)}\right),$$

and

$$\left(\frac{1}{k(N+k)}, \frac{1}{k(k_2+k)}\right).$$

Thus

$$Q_0(m) = \sum_{\substack{k=1 \\ (k,6)=6}}^N B_k(m) \int_{-\frac{1}{k(N+k)}}^{\frac{1}{k(N+k)}} \exp\left(\frac{\pi}{6k^2w} + \pi w(2m - \frac{1}{6})\right) d\phi,$$

where

$$(5.1) \quad B_k(m) = \sum'_{h \bmod k} \exp\left(-\frac{2\pi i h m}{k}\right) \Omega_{h,k} \text{ for } (k, 6) = 6.$$

Let R be the positive circuit of the rectangular path with vertices

$$\pm N^{-2} \pm \frac{i}{k(N+k)}.$$

Then, using $w = N^{-2} - i\phi$, the integral in $Q_0(m)$ above may be equated to

$$\begin{aligned} (5.2) \quad & \frac{1}{i} \int_R \exp\left(\frac{\pi}{6k^2w} + \pi w(2m - \frac{1}{6})\right) dw \\ & - \frac{1}{i} \int_{N^{-2} + \frac{i}{k(N+k)}}^{-N^{-2} + \frac{i}{k(N+k)}} - \frac{1}{i} \int_{-N^{-2} + \frac{i}{k(N+k)}}^{-N^{-2} - \frac{i}{k(N+k)}} - \frac{1}{i} \int_{-N^{-2} - \frac{i}{k(N+k)}}^{N^{-2} - \frac{i}{k(N+k)}} \\ & = 2\pi L_k(m) - \frac{1}{i} (J_1 + J_2 + J_3), \end{aligned}$$

where all four integrals have the same integrand.

The integral

$$L_k(m) = \frac{1}{2\pi i} \int_R \exp\left(\frac{\pi}{6k^2w} + \pi w(2m - \frac{1}{6})\right) dw$$

may be expressed in terms of well known integrals from the theory of Bessel

functions. For the Bessel function of the first kind with purely imaginary argument⁸ we have

$$(5.3) \quad I_\rho(z) = \frac{(z/2)^\rho}{2\pi i} \int_{-\infty}^{(0+)} t^{-\rho-1} \exp\left(t + \frac{z^2}{4t}\right) dt$$

and

$$(5.4) \quad I_{\rho-1}(z) - I_{\rho+1}(z) = \frac{2\rho}{z} I_\rho(z),$$

the latter being used with $\rho = 0$. From these we obtain

$$(5.5) \quad L_k(m) = \frac{1}{k\sqrt{12m-1}} I_1\left(\frac{\pi\sqrt{12m-1}}{3k}\right).$$

Now, along the paths of integration in J_1 and J_3 we have

$$w = u \pm \frac{i}{k(N+k)}, \quad -N^{-2} \leq u \leq N^{-2}.$$

It follows that

$$R(w) = u \leq N^{-2}, \quad R(1/w) = \frac{u}{u^2 + \frac{1}{k^2(N+k)^2}} < N^{-2}k^2(N+k)^2 \leq 4k^2$$

so that the absolute value of the integrand in each of J_1 and J_3 is

$$\leq \exp\left(\frac{2}{3}\pi + 2\pi mN^{-2}\right)$$

and hence

$$\left\{ \begin{array}{l} |J_1| \\ |J_3| \end{array} \right\} \leq 2N^{-2} \exp\left(\frac{2}{3}\pi + 2\pi mN^{-2}\right).$$

In J_2 we have $w = -N^{-2} + iv$, where

$$-\frac{1}{k(N+k)} \leq v \leq \frac{1}{k(N+k)}.$$

It follows that

$$R(w) = -N^{-2} < 0, \quad R(1/w) = \frac{-N^{-2}}{N^{-4} + v^2} < 0,$$

so that the absolute value of the integrand is less than unity, whence

$$|J_2| < \frac{2}{k(N+k)} < 2k^{-1}N^{-1}.$$

Collecting the last few results, and recalling that

$$B_k(m) = O(k^{2/3+\epsilon}m^{1/3}),$$

we obtain

$$(5.6) \quad Q_0(m) = 2\pi \sum_{\substack{k=1 \\ (k,6)=6}}^N B_k(m) L_k(m) + O(\exp(2\pi mN^{-2})m^{1/3}N^{-(1/3)+\epsilon}).$$

⁸ G. N. Watson, *Theory of Bessel Functions* (Cambridge, 1922), p. 181, (1), p. 79, (1).

We now treat $Q_1(m)$, and point out that $Q_2(m)$ may be handled analogously. We had

$$Q_1(m) = \sum_{\substack{k=1 \\ (k,6)=6}}^N \sum'_{h \bmod k} \exp\left(-\frac{2\pi i h m}{k}\right) \Omega_{h,k} \\ \times \sum_{l=k_1+k}^{N+k-1} \int_{-\frac{1}{kl}}^{-\frac{1}{k(l+1)}} \exp\left(\frac{\pi}{6k^2 w} + \pi w(2m - \frac{1}{6})\right) d\phi.$$

Interchanging the summations with respect to l and h , we obtain

$$(5.7) \quad Q_1(m) = \sum_{\substack{k=1 \\ (k,6)=6}}^N \sum_{l=k_1+k}^{N+k-1} \int_{-\frac{1}{kl}}^{-\frac{1}{k(l+1)}} \exp\left(\frac{\pi}{6k^2 w} + \pi w(2m - \frac{1}{6})\right) d\phi \\ \cdot \sum'_{\substack{h \bmod k \\ N < k+k_1 \leq l}} \exp\left(-\frac{2\pi i h m}{k}\right) \Omega_{h,k}.$$

The inner sum may be treated as was (4.8), yielding

$$O(k^{2/3+\epsilon} m^{1/3}).$$

Noting that

$$R\left(\frac{\pi}{6k^2 w}\right) = \frac{\pi N^{-2}}{6k^2(N^{-4} + \phi^2)} \\ \leq \frac{\pi}{6k^2 N^2(N^{-4} + k^{-2}(N+k)^{-2})} \leq \frac{\pi}{6(k^2 N^{-2} + \frac{1}{4})} < \pi$$

and

$$R(\pi w(2m - \frac{1}{6})) < 2\pi m N^{-2},$$

we conclude that

$$Q_1(m) = O\left(\sum_{k=1}^N \sum_{l=k_1+k}^{N+k-1} \left\{ \frac{1}{kl} - \frac{1}{k(l+1)} \right\} \exp(2\pi m N^{-2}) k^{2/3+\epsilon} m^{1/3}\right) \\ = O\left(\sum_{k=1}^N \frac{1}{kN} \exp(2\pi m N^{-2}) k^{2/3+\epsilon} m^{1/3}\right) \\ = O(N^{-(1/3)+\epsilon} \exp(2\pi m N^{-2}) m^{1/3}).$$

Since a similar result is valid for $Q_2(m)$, we combine this result with (4.13) and (5.6) to obtain

$$(5.8) \quad a_m^{(6)} = 2\pi \sum_{\substack{k=1 \\ (k,6)=6}}^N B_k(m) L_k(m) + O(\exp(2\pi m N^{-2}) m^{1/3} N^{-(1/3)+\epsilon}).$$

6. Estimations for $a_m^{(3)}$ and $a_m^{(2)}$. If the power series expansion for

$F^{-1}(x)$ is $\sum_{n=0}^{\infty} b_n x^n$ then $a_m^{(3)}$ is, from (3.3) and (2.11),

$$(6.1) \quad \sum_{\substack{k=1 \\ (k,6)=3}}^N \sum_{n=0}^{\infty} b_n \sum'_{h \bmod k} \exp \left\{ -\frac{\pi i}{k} (2hm - h'n) \right\} \Omega_{h,k} \\ \times \int_{-\theta'}^{\theta''} \exp \left\{ -\frac{\pi}{k^2 w} (n + \frac{1}{2}) + \pi w (2m - \frac{1}{6}) \right\} d\phi.$$

This differs from $a_m^{(6)}$ in that the coefficient of $1/w$ in the exponent of the integrand is always negative, whereas in $a_m^{(6)}$ it was positive for $n = 0$, which forced us to consider this case separately (section 5). It is clear, then, that we proceed with (6.1) as we did with I_2 in section 4. Formulas (2.4) and (2.11) imply

$$(6.2) \quad \Omega_{h,k} = -\exp \left\{ \frac{\pi i}{12} \left(\frac{2hk}{3} + \frac{h'k}{3} + \frac{2h}{k} - \frac{h'}{k} \right) \right\} \text{ for } (k, 6) = 3$$

provided that h' is chosen so that

$$hh' \equiv -1 \pmod{3k}.$$

Corresponding to (4.4) we now have

$$(6.3) \quad \sum'_{h \bmod k} \exp \left\{ -\frac{2\pi i}{3k} \left\{ h \left(3m - \frac{k^2+3}{12} \right) + \frac{h'}{2} \left(-3n - \frac{h^2+3}{12} \right) \right\} \right\}.$$

Since $(k, 6) = 3$, $(k^2 + 3)$ is divisible by 12. Also we may choose h' to be even, restricting it to the interval $0 < h' < 6k$. Then set $h'' = h'/2$, so that

$$0 < h'' < 3k \quad \text{and} \quad hh'' \equiv \frac{3k-1}{2} \pmod{3k}.$$

The sum (6.3) is then an incomplete Kloosterman sum mod $3k$, since for $hh'' \equiv a \pmod{k}$, $(a, k) = 1$

$$\sum'_{h \bmod k} \exp \left\{ \frac{2\pi i}{k} (uh + vh'') \right\} = \sum'_{h \bmod k} \exp \left\{ \frac{2\pi i}{k} (uh + avh') \right\}$$

where $hh' \equiv 1 \pmod{k}$. Corresponding to (4.10) we have

$$[6.4] \quad \sum'_{\substack{h \bmod k \\ N < k+h'' \leq l}} \exp \left\{ -\frac{2\pi i}{3k} \left\{ h \left(3m - \frac{k^2+3}{12} \right) + \frac{h'}{2} \left(-3n - \frac{k^2+3}{12} \right) \right\} \right\}$$

to which we apply the argument following (4.10). Thus

$$(6.5) \quad a_m^{(3)} = O(\exp(2\pi m N^{-2}) m^{1/3} N^{-(1/3)+\epsilon}).$$

The quantity

$$a_m^{(2)} = \sum_{\substack{k=1 \\ (k,6)=2}}^N \sum_{n=0}^{\infty} b_n \sum'_{h \bmod k} \exp \left\{ -\frac{2\pi i}{3k} (3hm - h'n) \right\} \Omega_{h,k} \\ \times \int_{-\theta'}^{\theta''} \exp \left\{ -\frac{\pi}{3k^2 w} (2n + \frac{1}{6}) + \pi w (2m - \frac{1}{6}) \right\} d\phi,$$

derived from formulas (2.8), (2.12), and (3.3), may be treated similarly. From (2.3) and (2.12) we obtain

$$\Omega_{h,k} = -\exp \left\{ \frac{\pi i}{12} \left(-2hk + \frac{h'k}{3} + \frac{2h}{k} + \frac{2h'}{3k} \right) \right\} \text{ for } (k, 6) = 2,$$

so that the Kloosterman sum is

$$\sum'_{\substack{h \bmod k \\ N < k+k_2 \leq t}} \exp \left\{ -\frac{2\pi i}{4k} \left\{ h \left(4m + \frac{k^2-1}{3} \right) + h''' \left(-4n - \frac{k^2+2}{6} \right) \right\} \right\}$$

where

$$(k, 6) = 2 \text{ and } hh' \equiv \frac{4k-1}{3} \text{ or } \frac{8k-1}{3} \pmod{4k}$$

according as $k \equiv 1$ or $2 \pmod{3}$. This sum is incomplete $\pmod{4k}$, and we conclude that

$$(6.6) \quad a_m^{(2)} = O(\exp(2\pi m N^{-2}) m^{1/2} N^{-(1/3)+\epsilon}).$$

7. The evaluation of $a_m^{(1)}$. The quantity

$$(7.1) \quad a_m^{(1)} = \sum_{\substack{k=1 \\ (k,6)=1}}^N \sum_{n=0}^{\infty} a_n \sum'_{h \bmod k} \exp \left\{ -\frac{2\pi i}{6k} (6hm - h'n) \right\} \Omega_{h,k} \\ \times \int_{-\theta'}^{\theta''} \exp \left\{ \frac{\pi}{36k^2w} (1 - 12n) + \pi w (2m - 1/6) \right\} d\phi$$

resembles $a_m^{(2)}$ in that the coefficient of $1/w$ in the exponent of the integrand is positive for $n=0$, so that the first term of the inner sum must be treated separately, which treatment follows that of section 5. Evaluating $\Omega_{h,k}$ with the aid of (2.4) and (2.13), we have

$$(7.2) \quad \Omega_{h,k} = \sum'_{h \bmod k} \exp \left\{ -\frac{\pi i}{k} \left(2hk - \frac{2h}{k} - \frac{h'k}{3} + \frac{h'}{3k} \right) \right\} \text{ for } (k, 6) = 1.$$

We may choose h' divisible by 6 and satisfying

$$hh' \equiv -1 \pmod{k}$$

since $(k, 6) = 1$. The Kloosterman sum is

$$(7.2) \quad \sum'_{h \bmod k} \exp \left\{ -\frac{2\pi i}{k} \left\{ h \left(m + \frac{k^2-1}{12} \right) + \frac{h'}{6} \left(n - \frac{k^2-1}{12} \right) \right\} \right\}.$$

Note that (k^2-1) is divisible by 12. Then $h'/6$ may be replaced by an integer h^{IV} satisfying

$$hh^{IV} \equiv \frac{k-1}{6} \text{ or } \frac{5k-1}{6} \pmod{k}$$

according as $k \equiv 1$ or $-1 \pmod{6}$. Thus we obtain a result similar to (5.8),

$$(7.3) \quad a_m^{(1)} = 2\pi \sum_{\substack{k=1 \\ (k,6)=1}}^N B_k(m) L_k(m) + O(\exp(2\pi m N^{-2}) m^{1/3} N^{-(1/3)+\epsilon}),$$

wherein

$$(7.4) \quad L_k(m) = \frac{1}{k\sqrt{72m-6}} I_1\left(\frac{\pi\sqrt{72m-6}}{18k}\right)$$

and $B_k(m)$ has the same form as in (5.1) but with $(k, 6)$ now equal to unity, $\Omega_{k,k}$ being defined by (7.2) in this case.

We now collect our results. Formulas (5.8), (6.5), (6.6), and (7.3) combine to give

$$(7.5) \quad a_m = a_m^{(6)} + a_m^{(3)} + a_m^{(2)} + a_m^{(1)} \\ = 2\pi \sum_{k=1}^N B_k(m) L_k(m) + O(\exp(2\pi m N^{-2}) m^{1/3} N^{-(1/3)+\epsilon})$$

provided we set

$$(7.6) \quad L_k(m) = 0 \text{ when } (k, 6) = 3 \text{ or } 2.$$

In (7.5) we hold m fixed and let N become infinite so that the error term becomes zero, and a_m is expressed by the convergent series

$$(7.7) \quad a_m = 2\pi \sum_{k=1}^{\infty} B_k(m) L_k(m)$$

the various quantities in this result being given by formulas (5.1) with (4.1) and (7.2), (5.5), (7.4), and (7.6).

UNIVERSITY OF PENNSYLVANIA.

FINITE METABELIAN GROUPS AND PLÜCKER LINE-COÖRDINATES.*

By H. R. BRAHANA.

1. Introduction. We are concerned with finite metabelian groups whose operators are all, except identity, of order p . The metabelian groups are those whose commutators are all in the central. If G is such a group and C is its central, then G/C is abelian and of type $1, 1, \dots$. Corresponding to every subgroup of G/C there is a subgroup of G which is either abelian or metabelian. Whenever G/C is the direct product of two of its subgroups, Γ_1 and Γ_2 , which are such that the corresponding subgroups G_1 and G_2 of G are both abelian, then G is a subgroup of the holomorph of G_1 and also a subgroup of the holomorph of G_2 . Conversely, when G is a subgroup of the holomorph of one of its abelian subgroups, then G/C is the direct product of two subgroups, Γ_1 and Γ_2 , which correspond to abelian subgroups of G . A method of classification of groups possessing this property has been given.¹ Our main interest here is in the groups which do not possess this property.

A group G which is not abelian has a commutator subgroup which is not identity. The central C either coincides with this commutator subgroup or else is the direct product of it and an abelian subgroup C' which is of type $1, 1, \dots$. In the latter case G itself is the direct product of C' and a metabelian group G' which has all the interesting properties of G .² We shall assume in what follows that the central and the commutator subgroup of G coincide.

The abelian subgroup C of G is not maximal abelian, for the group $\{C, s\}$, where s is any operator of G not in C , is abelian. The group $\{C, s\}$ may or may not be maximal abelian. Whether or not $\{C, s\}$ is maximal abelian will depend in general on the choice of s . The possession of a maximal abelian subgroup $\{C, s\}$ will be a characteristic property of G . We propose to investigate such groups G as have a maximal abelian subgroup $\{C, s\}$.³ This class

* Received December 12, 1938; Revised September 22, 1939.

¹ Cf. for references H. R. Brahana, "Metabelian groups and trilinear forms," *Duke Mathematical Journal*, vol. 1 (1935), pp. 185-197.

² More specifically, if two groups have the same order and possess the same G' , then they are simply isomorphic.

³ Cf. *American Journal of Mathematics*, vol. 56 (1934), p. 496. The theorem (5.2) which purports to deal with these groups is incorrect. We regret the incorrect theorem and the erroneous proof. The theorem was beside the main line of development of the paper and those which follow it.

of groups contains most of the groups which have been considered in the papers referred to above. However, our present investigation will be greatly facilitated by the work which has been done, for once it is recognized that G belongs to the holomorph of one of its abelian subgroups G is quickly identified by reference to those papers.

We denote the order of C by p^c ; we denote the maximal abelian subgroup $\{C, s\}$ of order p^{c+1} by H . Then we suppose G to be $\{H, U_1, U_2, \dots, U_k\}$. The order of G is p^{c+k} . If H is transformed by $U = \{U_1, U_2, \dots, U_k\}$ there is obtained a commutator subgroup, which lies in C . The fact that H is maximal abelian requires that the commutator subgroup obtained by transforming H by U be of order p^k . Hence we must have $c \geq k$. If U were abelian, G would be in the holomorph of H . We may therefore suppose that U is non-abelian. U then contains a commutator subgroup of order p^l where $l \geq 1$. If the commutators of pairs of U_i 's are all independent, then $l = k(k-1)/2$. Since C is the commutator subgroup as well as the central of G , we have $k \leq c \leq k(k+1)/2$. We note that G is generated by the k U_i 's and s . The numbers k and c are characteristic for a given G . We shall consider the groups G for a given k in subclasses according to the values of c .

When $c = k(k+1)/2$ the group is completely determined by the number k , since two such groups with the same k may be made simply isomorphic by letting generators correspond and letting commutators of corresponding pairs of generators correspond. We shall refer to this group as the *master group* for a given k . Such a group contains no abelian subgroup of order p^{c+2} . For all other values of c there exist groups G which contain abelian subgroups of order p^{c+2} . The possession of such subgroups is a characteristic property of a group and hence any set of invariants which determines the group must determine the number of such abelian subgroups contained in the group. Our method is to examine G for its abelian subgroups. This brings into prominence a certain matrix M whose properties give a set of definitive properties of G for $k < 4$. The elements of M are linear forms in certain indeterminates and the existence of certain abelian subgroups of G implies the existence of sets of values for the indeterminates which determine certain ranks for M . This investigation is carried out in § 2 for $k < 4$.

In § 3 the problem of the classification of these groups is approached from another direction, and for the case $k = 3$ is seen to be closely connected with Plücker line-coördinates in a finite three-space. The geometric formulation of the problem gives it an appearance of simplicity which would be misleading were we not warned by the intricacy of the considerations in § 2. The exposition of the close connection between these two seemingly distinct subjects is, of course, the important contribution of this paper.

2. The groups G for $k \leq 3$. When $k = 1$, then $c = 1$. G is of order p^3 and there is only one such metabelian group. This is the master group described in the preceding paragraph, and is otherwise well-known.⁴

When $k = 2$, then c is 2 or 3. There is but one group for $c = 3$ as was seen above. For $c = 2$ there is a group G which belongs to the holomorph of H . This group is generated by s , U_1 , and U_2 which satisfy the following relations with $\alpha = \beta = 0$:

$$(1) \quad \begin{aligned} U_1^{-1}sU_1 &= ss_1, & U_2^{-1}U_1U_2 &= U_1s_1^{\alpha}s_2^{\beta}, \\ U_2^{-1}sU_2 &= ss_2. \end{aligned}$$

$H = \{s, s_1, s_2\}$ is of order p^3 since s is not in C and H is maximal abelian.

Any group G with $k = 2$ is generated by operators satisfying (1) with α and β having suitable values. Two groups generated by operators satisfying (1) both having $\alpha = \beta = 0$ are obviously simply isomorphic. Hence for a group G , not in the holomorph of H , not both $\alpha = 0$ and $\beta = 0$. The subgroup of the holomorph of H described in the last paragraph contains the abelian subgroup $\{C, U_1, U_2\}$ whose order is $p^{c+2} = p^4$. If α and β exist, such that G contains no abelian subgroup of order p^4 , then the resulting G will be distinct from the one already obtained. An abelian subgroup of order p^4 will contain two independent operators which are not in C and neither is in the group generated by the other and C . Every operator of G can be written in the form $c_i s^a U_1^k U_2^l$, where c_i is some operator in C . The commutator of this and any other operator of G is independent of c_i ; hence for the purpose of investigating commutators we may assume that $c_i = 1$. Let $V_i = s^{a_i} U_1^{k_i} U_2^{l_i}$, $i = 1, 2$. Then

$$V_1^{-1}V_2^{-1}V_1V_2 = s_1^{a_1k_2 - a_2k_1} s_2^{a_1l_2 - a_2l_1} (s_1^{\alpha} s_2^{\beta})^{k_1l_2 - k_2l_1}.$$

If V_1 and V_2 are permutable this commutator is identity and we have the following congruences, mod p :

$$\begin{aligned} a_1k_2 - a_2k_1 + \alpha(k_1l_2 - k_2l_1) &\equiv 0, \\ a_1l_2 - a_2l_1 + \beta(k_1l_2 - k_2l_1) &\equiv 0. \end{aligned}$$

These are linear in a_2, k_2, l_2 . The matrix of coefficients is

$$M = \begin{pmatrix} -k_1 & a_1 - \alpha l_1 & \alpha k_1 \\ -l_1 & -\beta l_1 & a_1 + \beta k_1 \end{pmatrix},$$

whose rank is of course at most 2. Hence there is always a solution of the system of congruences. This corresponds to the fact that V_1 is permutable

⁴This is the only non-abelian group of order p^3 which contains operators of order p only. It is to be noted that $p \neq 2$, since the only groups whose operators are all except identity of order 2 are the abelian groups of type 1, 1, . . .

with any power of itself. If V_2 exists and is independent of and permutable with V_1 , then for some V_1 the above system must have two independent solutions and the rank of M must be at most 1. This requires that

$$\begin{aligned}\beta k_1 l_1 + a_1 l_1 - \alpha l_1^2 &\equiv 0, \\ \alpha k_1 l_1 - a_1 k_1 - \beta k_1^2 &\equiv 0.\end{aligned}$$

Since H is maximal abelian we cannot have $k_1 = l_1 = 0$, and hence the two conditions reduce to $\beta k_1 + a_1 - \alpha l_1 \equiv 0$. Therefore, whatever the values of α and β , a_1, k_1, l_1 may be found such that the rank of M is 1. Consequently, if $c = 2$, G contains an abelian subgroup of order p^{c+2} .

The abelian group $\{C, V_1, V_2\}$ does not contain s , for in that case H would not be maximal abelian. Hence G is generated by s, V_1 , and V_2 which satisfy the relations (1) with $\alpha = \beta = 0$. G is therefore a subgroup of the holomorph of H whenever $c = 2$.

When $k = 3$, we have $3 \leq c \leq 6$. There is one and only one group for $c = 6$. The simpler cases are those for which c is large; we consider the case where $c = 5$. The commutators obtained by transforming s by U_1, U_2 , and U_3 are independent and at least two of those obtained by transforming one U_i by another are independent of each other and the three preceding commutators. We may therefore suppose that generators of G satisfy the relations:

$$(2) \quad \begin{aligned}U_1^{-1} s U_1 &= s s_1, & U_2^{-1} U_1 U_2 &= U_1 s_1^{\alpha} s_2^{\beta} s_3^{\gamma} s_4^{\delta} s_5^{\epsilon}, \\ U_2^{-1} s U_2 &= s s_2, & U_3^{-1} U_1 U_3 &= U_1 s_4, \\ U_3^{-1} s U_3 &= s s_3, & U_3^{-1} U_2 U_3 &= U_2 s_5.\end{aligned}$$

There exists one such group with $\alpha = \beta = \gamma = \delta = \epsilon = 0$. This group contains the abelian subgroup $\{C, U_1, U_2\}$ whose order is p^{c+2} . Conversely, any group G with $k = 3$ and $c = 5$, which contains H as a maximal abelian subgroup and contains also an abelian subgroup of order p^{c+2} , is simply isomorphic with this group, for the abelian subgroup of order p^{c+2} contains two operators V_1 and V_2 such that $\{C, V_1, V_2\}$ is of order p^{c+2} and does not contain H . Then V_1 and V_2 may be used for U_1 and U_2 in relations (2).

Therefore, if any other group exists, it must contain no abelian subgroup of order p^{c+2} . Let $V_i = s^{a_i} U_1^{k_i} U_2^{l_i} U_3^{m_i}$, $i = 1, 2$. The condition that G have an abelian subgroup of order p^{c+2} is that there exist V_1 and V_2 which are independent and permutable. This leads as before to a set of congruences bilinear in the two sets of exponents. Considering these as linear congruences in the exponents of V_2 and writing the condition that the system have a solution, we obtain the matrix

$$M = \begin{pmatrix} -k_1 & a_1 - \alpha l_1 & \alpha k_1 & 0 \\ -l_1 & -\beta l_1 & a_1 + \beta k_1 & 0 \\ -m_1 & -\gamma l_1 & \gamma k_1 & a_1 \\ 0 & -m_1 - \delta l_1 & \delta k_1 & k_1 \\ 0 & -\epsilon l_1 & -m_1 + \epsilon k_1 & l_1 \end{pmatrix}.$$

If V_1 and V_2 independent and permutable exist it is necessary and sufficient that a_1, k_1, l_1, m_1 exist such that the rank of M is at most 2. Now a solution of the system of congruences is a set of values for a_2, k_2, l_2, m_2 which define a V_2 permutable with another operator and hence is a set of values which when used in place of a_1, k_1, l_1, m_1 in M reduce the rank to 2. Also the existence of such a V_2 implies that any operator of $\{V_1, V_2\}$ when expressed in terms of s, U_1, U_2, U_3 has a set of exponents which will reduce the rank of M to 2. Consequently, if there exists a V_1 which reduces the rank of M to 2, there exists a V_1 with $m_1 = 0$ which reduces the rank of M to 2. Letting $m_1 = 0$ and recalling that we may not have $a_1 = k_1 = l_1 = 0$ also, we obtain

$$\gamma + \beta\delta - \alpha\epsilon \equiv 0$$

as the condition that it be possible to reduce the rank of M to 2. Since there exist numbers $\alpha, \beta, \gamma, \delta, \epsilon$ which do not satisfy this condition there exist groups G with $c = 5$ which contain no abelian subgroup of order p^{c+2} .

We now determine a canonical form for a set of generating relations for the group with $c = 5$ and no abelian subgroup of order p^{c+2} , and in so doing show that these conditions are sufficient to determine the group. This canonical form is a particular set of values for $\alpha, \beta, \gamma, \delta, \epsilon$. This set of numbers determines the expression of the commutator of U_1 and U_2 in terms of the commutators of the other pairs of generators. Taking s_1, s_2, \dots, s_5 as defined by (2), we see that the commutator of U_1 and U_2 cannot be in $\{s_4, s_5\}$ for then the group $\{U_1, U_2, U_3\}$ would be that metabelian group we considered with $k = 2$ and $c = 2$ and hence would contain an abelian subgroup of order p^4 . Though this group is of order p^4 , it would determine an abelian subgroup of order p^{c+2} of G , namely, the direct product of the group of order p^4 and $\{s_1, s_2, s_3\}$. Therefore the commutator subgroup of $\{U_1, U_2, U_3\}$ has a subgroup of order p in common with $\{s_1, s_2, s_3\}$. Every operator of the commutator subgroup of $\{U_1, U_2, U_3\}$ is a commutator, for otherwise some quotient group of $\{U_1, U_2, U_3\}$ would be a metabelian group with $k = 2, c = 2$, and no abelian subgroup of order p^4 . We have seen that no such group exists. Therefore, U'_1 and U'_2 may be chosen in U so that their commutator is in $\{s_1, s_2, s_3\}$. Let s'_i be the commutator of U'_i and s . The commutator of U'_1 and U'_2 cannot be in $\{s'_1, s'_2\}$; for then $\{H, U'_1, U'_2\}$ would have a commu-

tator subgroup of order p^2 and hence would have an abelian subgroup of order p^{c+2} . We may therefore denote the commutator of U'_1 and U'_2 by s'_3 , and then find in U an operator U'_3 independent of U'_1 and U'_2 which with s has s'_3 for a commutator. Therefore, any group having the given properties has generators which satisfy (2) with $\alpha = \beta = \delta = \epsilon = 0$ and $\gamma = 1$. Thus there are just two groups with $k = 3$ and $c = 5$, and they are distinguished by the fact that one contains an abelian subgroup of order p^{c+2} and the other does not.

When $c = 4$ the considerations which we have employed above show that G is generated by operators satisfying the relations:

$$(3) \quad \begin{aligned} U_1^{-1}sU_1 &= ss_1, & U_2^{-1}U_1U_2 &= U_1s_1\alpha_1s_2\beta_1s_3\gamma_1, \\ U_2^{-1}sU_2 &= ss_2, & U_3^{-1}U_1U_3 &= U_1s_1\alpha_2s_2\beta_2s_3\gamma_2, \\ U_3^{-1}sU_3 &= ss_3, & U_3^{-1}U_2U_3 &= U_2s_4. \end{aligned}$$

The condition for permutability of V_1 and V_2 is that the matrix

$$M = \begin{pmatrix} -k_1 & a_1 - \alpha_1l_1 - \alpha_2m_1 & \alpha_1k_1 & \alpha_2k_1 \\ -l_1 & -\beta_1l_1 - \beta_2m_1 & a_1 + \beta_1k_1 & \beta_2k_1 \\ -m_1 & -\gamma_1l_1 - \gamma_2m_1 & \gamma_1k_1 & a_1 + \gamma_2k_1 \\ 0 & 0 & -m_1 & l_1 \end{pmatrix}$$

be of rank 2. Again, if V_1 exists such that the rank of M is 2 then there exists such a V_1 with $m_1 = 0$. In order that M be of rank 2 when $m_1 = 0$ it is necessary that $\gamma_1 = 0$ or else that $l_1 = 0$. The latter possibility requires further that

$$a_1^2 + (\beta_1 + \gamma_2)a_1k_1 + (\beta_1\gamma_2 - \beta_2\gamma_1)k_1^2 \equiv 0.$$

In order that a_1 and k_1 , rational and not both zero, exist and satisfy this relation it is necessary and sufficient that

$$(4) \quad (\beta_1 - \gamma_2)^2 + 4\beta_2\gamma_1$$

be a square, mod p . Since it is possible to select numbers $\alpha_1, \beta_1, \dots, \gamma_2$ with $\gamma_1 \neq 0$ so that this condition is not satisfied it follows that there exist groups G with $c = 4$ which contain no abelian subgroup of order p^{c+2} . When $\gamma_1 = 0$, then (4) is a square, and therefore, that (4) be a square, is a necessary as well as a sufficient condition for G to contain an abelian subgroup of order p^{c+2} .

Two groups which have $k = 3$, $c = 4$, and which contain no abelian subgroup of order p^{c+2} are simply isomorphic. One such group is generated by operators satisfying (3) with $\alpha_1 = \beta_1 = \alpha_2 = \gamma_2 = 0$, $\beta_2 = 1$, $\gamma_1 = r$, where r is a particular not-square. New generators U'_1, U'_2, U'_3 may be selected in U so that

$$(5) \quad \begin{aligned} c'_{12} &= s'_1 a_1 s'_2 \beta_1 s'_3 \gamma_1 \\ c'_{13} &= s'_1 a_2 s'_2 \beta_2 s'_3 \gamma_2, \end{aligned}$$

where c'_{ij} is the commutator of U'_i and U'_j , s'_i is the commutator of U'_i and s , and $\alpha_1, \beta_1, \dots, \gamma_2$ are arbitrary except for the condition that (4) be not a square. Let

$$\begin{aligned} U'_1 &= U_1, \\ U'_2 &= U_1^{k_1} U_2^{l_1} U_3^{m_1}, \\ U'_3 &= U_1^{k_2} U_2^{l_2} U_3^{m_2}. \end{aligned}$$

If the commutators are obtained and conditions derived that they satisfy (5), there are obtained six linear non-homogeneous congruences in the six unknowns k_1, \dots, m_2 . The ranks of the matrix of coefficients and the augmented matrix are the same provided $\beta_1 \gamma_2 - \beta_2 \gamma_1 \neq 0$, which is true whenever (4) is not a square. This completes the proof of the statement at the beginning of the paragraph.

For any other group whose generators satisfy (3) we may suppose that (4) is a square. Then there exists a set $a_1, k_1, 0, 0$ which reduces the rank of M to at most 2. If for a particular set the rank of M becomes 1, then the corresponding system of congruences has three linearly independent solutions. These solutions define V_1 itself and two others which we denote by V_2 and V_3 . Then every element of $\{V_2, V_3\}$ is permutable with V_1 . Hence G contains at least $p + 1$ abelian subgroups of order p^{c+2} . Two such groups are simply isomorphic since V_1, V_2, V_3 may be taken as generators. They satisfy (3) with $\alpha_1 = \beta_1 = \dots = \gamma_2 = 0$. Conditions that the rank of M may be made 1 are: $\gamma_2 = \beta_1, \beta_2 = \gamma_1 = 0$, since $k_1 = 0$ implies $a_1 = 0$.

For the remaining groups we may suppose that M becomes of rank 2 for $l_1 = m_1 = 0$ and a_1 and k_1 satisfying the quadratic which precedes (4). Writing this quadratic in the form

$$(a_1 - \lambda_1 k_1)(a_1 - \lambda_2 k_1) \equiv 0,$$

we note that $a_1/k_1 = \lambda_1$ or λ_2 and unless $\lambda_1 = \lambda_2$ there exist two independent sets $a_1, k_1, 0, 0$ and $a'_1, k'_1, 0, 0$ each of which reduces the rank of M to 2. When $\lambda_1 \neq \lambda_2$ the system of congruences for the determination of a_2, k_2, l_2, m_2 becomes

$$\begin{aligned} -a_2 + \lambda_i k_2 + \alpha_1 l_2 + \alpha_2 m_2 &\equiv 0, \\ (\lambda_i - \beta_1) l_2 - \beta_2 m_2 &\equiv 0, \\ \gamma_1 l_2 + (\lambda_i - \gamma_2) m_2 &\equiv 0, \end{aligned}$$

where the last two are dependent. The following sets reduce M to a matrix of rank 2:

$$\begin{array}{ll} \lambda_1, 1, 0, 0 & a_2, k_2, -\beta_2, \lambda_1 + \beta_1 \\ \lambda_2, 1, 0, 0 & a'_2, k'_2, -\beta_2, \lambda_2 + \beta_1. \end{array}$$

The sets in the same row determine permutable operators V_1, V_2 and V'_1, V'_2 . If $\beta_2 \neq 0$ and $\lambda_1 \neq \lambda_2$, then these four operators are independent and they generate G . G then contains two abelian subgroups of order p^{c+2} and is a subgroup of the holomorph of $\{C, V_1, V_2\}$. Such a group is completely determined by the above mentioned properties.⁵ It is not simply isomorphic with any of the groups determined previously. The assumption that $\beta_2 \neq 0$ is not essential, for if $\beta_2 = 0$, we solve the first and third congruences and obtain the same result so far as abelian subgroups of order p^{c+2} are concerned and these abelian subgroups determine G .

If $\lambda_1 = \lambda_2$ then considerations similar to those used above show that G contains but one abelian subgroup of order p^{c+2} . Under the given conditions on the exponents $\alpha_1, \dots, \gamma_2$ new generators U'_1, U'_2, U'_3 in U may be found such that $\beta'_2 = 1$ and $\alpha'_1 = \beta'_1 = \gamma'_1 = \alpha'_2 = \gamma'_2 = 0$. This shows the existence and the uniqueness of the group with one abelian subgroup of order p^{c+2} .

Hence for $k = 3$ and $c = 4$ there are four groups having respectively 0, 1, 2, and $p + 1$ abelian subgroups of order p^{c+2} ; any two groups with the same number of abelian subgroups of order p^{c+2} are simply isomorphic.

For $c = 3$ generators of G satisfy the relations:

$$(6) \quad \begin{array}{ll} U_1^{-1}sU_1 = ss_1, & U_2^{-1}U_1U_2 = U_1s_1\alpha_1s_2\beta_1s_3\gamma_1, \\ U_2^{-1}sU_2 = ss_2, & U_3^{-1}U_1U_3 = U_1s_1\alpha_2s_2\beta_2s_3\gamma_2, \\ U_3^{-1}sU_3 = ss_3, & U_3^{-1}U_2U_3 = U_2s_1\alpha_3s_2\beta_3s_3\gamma_3. \end{array}$$

Conditions for the permutability of V_1 and V_2 require the rank of

$$M = \begin{pmatrix} -k_1 & a_1 - \alpha_1l_1 - \alpha_2m_1 & \alpha_1k_1 - \alpha_3m_1 & \alpha_2k_1 + \alpha_3l_1 \\ -l_1 & -\beta_1l_1 - \beta_2m_1 & a_1 + \beta_1k_1 - \beta_3m_1 & \beta_2k_1 + \beta_3l_1 \\ -m_1 & -\gamma_1l_1 - \gamma_2m_1 & \gamma_1k_1 - \gamma_3m_1 & a_1 + \gamma_2k_1 + \gamma_3l_1 \end{pmatrix}$$

to be at most 2 for a proper choice of V_1 . For certain groups G , in other words for certain sets α, β, γ , it is possible to choose V_1 so that M has rank 1. In such a case V_1 is permutable with V_2 and V_3 and V_1, V_2, V_3 are independent. Since H is maximal abelian, s is not contained in $\{V_1, V_2, V_3\}$, each operator of which reduces the rank of M to 2 or 1. Hence G is generated by s, V_1, V_2 , and V_3 . We may take U_i to be V_i and assume G to be generated by operators which satisfy (6) with $\alpha_1 = \beta_1 = \gamma_1 = \alpha_2 = \beta_2 = \gamma_2 = 0$. If in addition $\alpha_3 = \beta_3 = \gamma_3 = 0$, then U_2 and U_3 also reduce the rank of M to 1. This group is generated by the two abelian subgroups H and U , and is

⁵ Cf. "On the metabelian groups which contain a given group H as a maximal invariant abelian subgroup," *American Journal of Mathematics*, vol. 56 (1934), p. 510. This is the first group in the table.

therefore a subgroup of the holomorph of H . It is completely determined by its order and the order of its commutator subgroup.⁶ It contains $p^2 + p + 1$ abelian subgroups of order p^{c+2} corresponding to the same number of subgroups of order p^2 in U .

If $\alpha_3, \beta_3, \gamma_3$ are not all zero, we note that every element of $\{U_2, U_3\}$ reduces the rank of M to 2, since every such element is permutable with U_1 . G therefore contains at least $p + 1$ abelian subgroups of order p^{c+2} , one for each subgroup of order p in $\{U_2, U_3\}$. It will be convenient to write M in the more special form

$$M' = \begin{pmatrix} -k_1 & a_1 & -\alpha_3 m_1 & \alpha_3 l_1 \\ -l_1 & 0 & a_1 - \beta_3 m_1 & \beta_3 l_1 \\ -m_1 & 0 & -\gamma_3 m_1 & a_1 + \gamma_3 l_1 \end{pmatrix}.$$

The choice $a_1, k_1, l_1, m_1 = 0, \alpha_3, \beta_3, \gamma_3$ reduces the rank of M' to 1. The corresponding operator V_1 is permutable with U_1 since every operator of U is permutable with U_1 ; it is permutable with V_2 determined by $\beta_3, 0, 0, 1$ which is not in U if $\beta_3 \neq 0$; and it is permutable with V_3 determined by $-\gamma_3, 0, 1, 0$ which is not in U if $\gamma_3 \neq 0$. If neither β_3 nor γ_3 is zero, the operator V_3 is in the group $\{U_1, V_1, V_2\}$ since

$$\begin{array}{cccc} 0, & \alpha_3, & \beta_3, & \gamma_3 \\ \beta_3, & 0, & 0, & 1 \\ -\gamma_3, & 0, & 1, & 0 \\ 0, & 1, & 0, & 0 \end{array}$$

are linearly dependent. If not both β_3 and γ_3 are zero, then G is generated by U_1, U_2, V_1 , and V_2 or V_3 . The pairs U_1, U_2 and V_1, V_2 (or V_1, V_3) are permutable. Hence if not both β_3 and γ_3 are zero G is generated by two of its abelian subgroups and hence is a subgroup of the holomorph of $\{C, U_1, U_2\}$. This group is generated by operators which satisfy the relations

$$\begin{aligned} U_1^{-1} s_1 U_1 &= s_1 s_3, & U_2^{-1} s_1 U_2 &= s_1 s_5, \\ U_1^{-1} s_2 U_1 &= s_2 s_4. \end{aligned}$$

The group is unique,⁷ it contains $2p + 1$ abelian subgroups of order p^{c+2} .

If $\beta_3 = \gamma_3 = 0$ and $\alpha_3 \neq 0$, then the rank of M' is 3 unless $a_1 = 0$; and if $a_1 = 0$ the rank is 2 unless $l_1 = m_1 = 0$. Therefore U_1 is the only operator of G which reduces the rank of M' to 1. The only abelian subgroups of order p^{c+2} of G correspond to subgroups of order p^2 of U which contain U_1 . Hence

⁶ Cf. the preceding reference, p. 495, Theorem 5.1.

⁷ Cf. *loc. cit.*, p. 510. This is the group of order p^{n+2} with K of order p^3 and one subgroup of type 1.

G contains $p + 1$ such subgroups which distinguishes G from the groups which have been discussed. These $p + 1$ abelian subgroups of order p^{c+2} are all contained in a non-abelian subgroup of order p^{c+3} , generated by C , U_1 , U_2 , and U_3 , which will distinguish it from another group containing $p + 1$ abelian subgroups of order p^{c+2} which will follow. In the present case an obvious change of generators will make $\alpha_3 = 1$, which shows that any two such groups are simply isomorphic. G is not in the holomorph of any of its abelian subgroups since all of its abelian subgroups are in H or in $\{C, U\}$ and none of the latter contains both U_2 and U_3 .

For none of the remaining groups with $c = 3$ can V_1 be selected to make the rank of M smaller than 2. We consider those for which the rank of M may be reduced to 2. For these we may assume $\alpha_1 = \beta_1 = \gamma_1 = 0$. Then both U_1 and U_2 reduce the rank of M to 2. If there exists a V_1 not in $\{U_1, U_2\}$ which also reduces the rank to 2, the corresponding V_2 will not be in $\{U_1, U_2\}$ for otherwise V_2 would reduce the rank to 1. Every operator of $\{V_1, V_2\}$ will reduce the rank to 2. Hence there exists a V_1 with $m_1 = 0$ and not in $\{U_1, U_2\}$ which reduces the rank to 2. Under these conditions M takes the form:

$$M'' = \begin{pmatrix} -k_1 & a_1 & 0 & \alpha_2 k_1 + \alpha_3 l_1 \\ -l_1 & 0 & a_1 & \beta_2 k_1 + \beta_3 l_1 \\ -0 & 0 & 0 & a_1 + \gamma_2 k_1 + \gamma_3 l_1 \end{pmatrix}.$$

Now V_1 , being not in $\{U_1, U_2\}$, must have $a_1 \neq 0$. Such a V_1 exists only if not both γ_2 and γ_3 are zero. Hence if $\gamma_2 = \gamma_3 = 0$, G contains but one abelian subgroup of order p^{c+2} . The condition that no operator of $\{U_1, U_2\}$ reduces the rank to 1 requires that $(\alpha_2 \beta_3 - \alpha_3 \beta_2)$ be different from zero which implies that $(\beta_3 - \alpha_2)^2 + 4\alpha_3 \beta_2$ be not a square. The existence of such a group is obvious; we omit for the moment consideration of the question of uniqueness. If not both γ_2 and γ_3 are zero, we may suppose that $\gamma_2 \neq 0$. Then $-\gamma_2, 1, 0, 0$ determines an operator V_1 which reduces the rank of M'' to 2. In this case G contains at least two abelian subgroups of order p^{c+2} , and since the rank of M cannot be made 1 the corresponding V_2 is not in $\{U_1, U_2, V_1\}$. Hence such a group, if it exists, is generated by the two abelian subgroups $\{C, U_1, U_2\}$ and $\{C, V_1, V_2\}$. It is therefore a subgroup of the holomorph of either. It is identified as the group⁸ with commutator subgroup of order p^3 and no subgroup of Type 1. The existence is established by showing that the group described in the paper referred to contains a maximal abelian subgroup of order p^{c+1} . It contains $p + 1$ abelian subgroups of order p^{c+2} . Since it is generated by two of these abelian subgroups it is distinguished from the group

⁸ Cf. *loc. cit.*

with $p + 1$ abelian subgroups of order p^{c+2} all contained in a subgroup of order p^{c+3} .

We shall now see that the next to the last group is uniquely determined by the property of having one abelian subgroup of order p^{c+2} . This group is obviously not in the holomorph of any of its abelian subgroups, since it contains no abelian subgroup of order p^{c+3} and but one of order p^{c+2} . We then fix attention on the characteristic subgroup $\{C, U_1, U_2\}$ which we denote by H' . With this change of notation

$$s'_1 = U_1, \quad s'_2 = U_2, \quad U'_1 = s_1^{-1}, \quad U'_2 = U_3^{-1}, \quad s'_3 = s_1, \quad s'_4 = s_2, \quad s'_5 = s_3$$

we have the following relations satisfied:

$$\begin{aligned} U'_1{}^{-1}s'_1U'_1 &= s'_1s'_3, & U'_2{}^{-1}s'_1U'_2 &= s'_1(s'_3{}^{\alpha_2}s'_4{}^{\beta_2})^{-1}, \\ U'_1{}^{-1}s'_2U'_1 &= s'_2s'_4, & U'_2{}^{-1}s'_2U'_2 &= s'_2(s'_3{}^{\alpha_3}s'_4{}^{\beta_3})^{-1}, \\ & & U'_2{}^{-1}U'_1U'_2 &= U'_1s'_5. \end{aligned}$$

The commutator subgroup arising from transformation of H' by $\{U'_1, U'_2\}$ is of order p^2 . U'_1 and U'_2 determine two permutable operators in the group of isomorphisms of H' which with H' give the particular subgroup⁹ of the holomorph of H' which has no subgroup of Type 1. A choice of generators to give the canonical form of generating relations of this subgroup of the holomorph of H' gives a canonical form for generating relations of G . This form is the above set with $\alpha_2 = \beta_3 = 0$, $\beta_2 = -1$, $\alpha_3 = -r$, where r is a particular not-square. The possibility of doing this is a consequence of the fact that $(\beta_3 - \alpha_2)^2 + 4\alpha_3\beta_2$ is not a square. The canonical form contains no arbitrary constants and hence the group is uniquely defined by the fact that it contains just one abelian subgroup of order p^{c+2} .

For arbitrary α, β, γ there exists V_1 such that the rank of M reduces to 2, when the corresponding a_i, k_i, l_i, m_i are substituted. A proof of this will establish the fact that every group G with $c = 3$ contains at least one abelian subgroup of order p^{c+2} , and hence is one of the five groups determined above.

If there exists a V_1 which reduces the rank of M to 2, there exists one with $m_1 = 0$. Suppose that $a_1 = 0$ also for this V_1 . The condition that such a V_1 exist is that $F(k_1, l_1) \equiv 0$, where F is

$$(\beta_1\gamma_2 - \beta_2\gamma_1)k_1^2 + (\beta_1\gamma_3 - \beta_3\gamma_1 - \alpha_1\gamma_2 + \alpha_2\gamma_1)k_1l_1 + (\alpha_3\gamma_1 - \alpha_1\gamma_3)l_1^2.$$

There exist quantities α, β, γ such that this congruence is irreducible; we may suppose that such is the case, for otherwise we have the existence of the required V_1 . Hence we may assume that $a_1 \neq 0$. When $a_1 \neq 0$, the first column of M is expressible linearly in terms of the last three columns, and hence the rank

⁹ Cf. *loc. cit.*, p. 510 and p. 500.

of M is the same as the rank of the matrix composed of the last three columns. The determinant of this matrix, with $m_1 = 0$, is

$$a_1[a_1^2 + \{(\beta_1 + \gamma_2)k_1 + (\gamma_3 - \alpha_1)l_1\}a_1 + F(k_1, l_1)].$$

The quadratic factor can always be made zero by a proper choice of a_1 , k_1 , l_1 , and since $F(k_1, l_1)$ has no linear factors a_1 will not be zero. This completes the proof.

We give below a table of the groups for $k < 4$. For given k and c the different groups are arranged in the order of their appearance in this paper.

k	c	number of abelian subgroups of order p^{c+2} .
1	1	0
2	3	0
	2	1
3	6	0
	5	1, 0
	4	0, $p + 1$, 2, 1
	3	$p^2 + p + 1$, $2p + 1$, $p + 1^*$, 1 , $p + 1^*$

* These two groups are distinguished by the fact that the first contains an abelian subgroup of order p^{c+3} and the other does not.

3. A geometric description of the groups for $k = 3$. It has been convenient in the preceding pages to single out the maximal abelian subgroup H , and consequently to distinguish between s and the other generators. In the sequel we shall drop this distinction and consider the same groups generated by the operators U_1, U_2, U_3 , and U_4 . Every such group is defined by a set of relations on the U_i 's. Each set contains six relations which define commutators of pairs of U_i 's. If there are no more relations, aside from those expressing permutability of the commutators, that is, if the six commutators are independent, then the group G is the master group described in Section 1. Any other group generated by four U_i 's is defined by additional relations among these six commutators and hence is a quotient group of this group of order p^{10} with respect to some subgroup of the commutator subgroup. The existence of two kinds of group with $c = 5$ shows that the commutator subgroup of G , of order p^6 , contains two kinds of cyclic subgroup. Distinguishing properties of these two groups with $c = 5$ are that one contains an abelian subgroup of order p^{c+2} and the other does not. G itself contains no abelian subgroup of order p^{c+2} and consequently if the group with $c = 5$ contains such a subgroup the process of taking the quotient group introduces permutability among operators of the form $U_1^{x_1}U_2^{x_2}U_3^{x_3}U_4^{x_4}$ where none existed in G .

This means that the cyclic group which is set equal to identity contains commutators. In the other case for $c = 5$ the cyclic group contains no commutator except identity. We therefore examine the question of commutators and non-commutators in the commutator subgroup of the master group G .

The group G is defined by the relations:

$$\begin{aligned} U_j^{-1} U_i U_j &= U_i r_{ij}, & (i < j, j = 2, 3, 4), \\ r_{ij} r_{kl} &= r_{kl} r_{ij}, & (i, j, k, l = 1, 2, 3, 4). \end{aligned}$$

If the r_{ij} 's are ordered then every operator of the commutator subgroup of G is determined by the set of exponents of the r_{ij} 's in the expression for it; two operators belong to the same cyclic group if and only if their sets of exponents are linearly dependent. The set $(0, 0, \dots, 0)$ corresponds to identity. Hence, if the sets of exponents are taken to be the coördinates of points in a finite projective space R of 5 dimensions, every point in R will correspond to a cyclic subgroup of the commutator subgroup of G . We determine the condition that the point $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_6)$ which corresponds to $r_{12}^{\alpha_1} r_{13}^{\alpha_2} r_{14}^{\alpha_3} r_{23}^{\alpha_4} r_{24}^{\alpha_5} r_{34}^{\alpha_6}$, represent a cyclic subgroup which contains a commutator.¹⁰ If α represents a commutator then there exist two operators

$$\begin{aligned} V_1 &= U_1^{x_1} U_2^{x_2} U_3^{x_3} U_4^{x_4} \\ V_2 &= U_1^{y_1} U_2^{y_2} U_3^{y_3} U_4^{y_4} \end{aligned}$$

which have the element corresponding to α for their commutator.

$$V_1^{-1} V_2^{-1} V_1 V_2 = r_{12}^{x_1 y_2 - x_2 y_1} r_{13}^{x_1 y_3 - x_3 y_1} \dots r_{34}^{x_3 y_4 - x_4 y_3}.$$

This is exactly the problem of the Plücker line coördinates in a projective three-space. We then have the following theorem:

A point α in the space R corresponds to a commutator if and only if it lies on the four-dimensional spread S defined by

$$\alpha_1 \alpha_6 - \alpha_2 \alpha_5 + \alpha_3 \alpha_4 \equiv 0.$$

A point in the space R corresponds to a cyclic subgroup in the commutator subgroup K of G ; a line in R , being the set of points linearly dependent on two points, corresponds to a subgroup of order p^2 in K ; and a plane corresponds to a subgroup of order p^3 in K . The effect of taking the elements of a particular subgroup of K to be identity, so far as abelian subgroups of order p^{c+2} of the resulting quotient group are concerned, depends on the relation of the corresponding point, line, or plane to the quadratic spread S . If the point, line, or plane has a point in common with S , the resulting quotient group will

¹⁰ Since $V_2^{-1} V_1^k V_2 = V_1^k r^k$, if r is the commutator of V_1 and V_2 , then every element of a cyclic group is a commutator or else none (except identity) is a commutator.

contain two permutable elements V_1 and V_2 which with C give an abelian subgroup of order p^{c+2} .

The points of R constitute two classes with respect to S , a point being either on S or not on S . If the point is not on S the corresponding quotient group contains no abelian subgroup of order p^{c+2} . The fact that there is but one such group for $c=5$ means that all points not on S are alike. The canonical form obtained for generating relations of this group in Section 2 has the commutator of U_1 and U_2 the same as that of U_3 and U_4 . Thus by putting equal to identity the elements of a cyclic group corresponding to a point not on S , a group of order p^2 determined by two points on S is reduced to a cyclic group of order p . Hence every point of R not on S is on a line joining two points of S .

A line in R may have no points on S ; it may have one point on S ; it may have two points on S ; or it may lie wholly on S . These possibilities correspond to the groups with $c=4$ with $0, 1, 2, p+1$ abelian subgroups of order p^{c+2} .

A plane in R has at least one point in common with S . It may cut S in one point, in a line, in a proper conic, in two lines, or it may lie wholly in S . These possibilities correspond to the groups with $c=3$ and $1, p+1, p+1, 2p+1$, and p^2+p+1 abelian subgroups of order p^{c+2} respectively.

Each of the last three groups is a subgroup of the holomorph of one of its abelian subgroups. A geometric criterion for this possibility involves a consideration of the three-space (x_1, x_2, x_3, x_4) . Let us consider the case $c=3$. A particular group is determined by a plane in R which has a certain set of points on S . Each point on S determines a line in X . Each line in X is determined by two of its points. Two skew lines in X will be determined by four points in terms of which every point of X can be expressed linearly. If two lines in X are not skew, then the line joining the corresponding points in R lies wholly on S .¹¹ Thus in the case of the third and fourth groups above, where the plane cuts S in a proper conic and in a conic consisting of a pair of lines, it is possible to select two points on the intersection which represent skew lines in X . In those two cases the groups are subgroups of the holomorph of the abelian group of order p^{c+2} . For the first two cases it is not possible to make such a selection. A plane wholly on S determines a set of lines in X which are also on a plane. The corresponding points on X represent an abelian group of order p^{c+3} in G . Any metabelian group with operators all of order p which contains an abelian subgroup of index p is in the holo-

¹¹ Cf. for example, Veblen and Young, *Projective Geometry*, vol. I (1910), p. 329, Theorem 30.

morph of that abelian subgroup. For $c = 4$ the group is determined by a line in R . The only one of these groups which is in the holomorph of one of its abelian subgroups corresponds to a line in R which has two and only two points on S . In the case where the line is wholly on S each of its points determines a line in X but all of these lines in X are in a plane. Since not every line in this plane in X is represented by a point on the intersection of S with the given line in R , this plane in X corresponds to a non-abelian subgroup of order p^{c+3} in the given group.

In general, for an arbitrary number k of generators, the condition that a group G' be in the holomorph of one of its abelian subgroups may be stated in geometric form. If G' is in the holomorph of one of its abelian subgroups, then the points of the space X can be expressed as linear combinations of points of two of its subspaces each of which determines an abelian subgroup of G' . An abelian subgroup of order p^{c+k_1} in G' determines a $(k_1 - 1)$ -space in X all of whose lines determine points on the intersection of S with the m -space in R which determines G' as a quotient group of the master group G .

The geometric aspect of the solution of the problem of classification of metabelian groups with k independent generators and elements all of order p is now clear. It involves the extension of the theory of Plücker line coördinates to a space of $k - 1$ dimensions. These coördinates determine a space R of $\gamma - 1$ dimensions, $\gamma = k(k - 1)/2$. In R the points which correspond to commutators are on a subspace S of $2k - 4$ dimensions defined by $(k - 2)(k - 3)/2$ quadratic congruences, the conditions that a point of R represent a line of X . It then involves the determination of the relations of points, lines, planes, three-spaces, etc. in R to the subspace S . These relations determine the possible types of quotient groups of G . In determining the types of relations to S of flat m -spaces in R it is necessary to separate the m -spaces of R into classes, all the members of a class being conjugate under the group of collineations of R which leave S invariant.¹² This group is closely connected with the group of collineations in X . Of course the transformations are "rational" and the geometry is finite.

¹² It is this part of the problem that made necessary most of detail of section 2.

AN EXTENSION OF ANALYTIC FUNCTIONS TO MATRICES.*

By R. W. WAGNER.

The analytic functions of a complex variable have many interesting properties. The property of analytic continuation is made the basis for this extension of such functions to matrices. The extended function is then a mapping of a subset of the matrix space on to another subset of the same space. The procedure followed here is to replace the complex variable in a power series by a variable matrix, show that the resulting matrix function can be reduced to the original function of several complex variables, and apply the process of analytic continuation to each variable. The most interesting results of this paper concern the singularities of the extended function which are introduced by the extension. The last part of the paper shows how this approach may be applied to the solution of certain matrix equations.

The notational scheme is as follows: Capital script letters indicate matrices, small letters indicate ordinary complex numbers, and subscripts are used for enumerative purposes.

1. Let \mathcal{M} denote the matrix space, the space of all square matrices of n rows whose elements are complex numbers. One can make this a metric space by defining the absolute value of a matrix and then defining the distance from X to Y to be the absolute value of $X - Y$. The absolute value of X will be taken to be¹

$$|X| = \sqrt{\text{tr } X\bar{X}'} = \sum_{i,j} x_{ij}\bar{x}_{ij}.$$

A similarity transformation applied to \mathcal{M} is just a change of coördinates in \mathcal{M} . Unfortunately, such a transformation does not leave the distance invariant, but it is a homeomorphic transformation of \mathcal{M} . Therefore, limiting relations will be independent of the coördinate system, and it will be permissible to use the most convenient coördinate system for investigating these limits.

It is convenient to distinguish several subsets of \mathcal{M} for future reference.

I. $\mathcal{R}(x)$, the set of matrices which have x for a characteristic root.

* Received November 28, 1938; Revised August 19, 1939.

¹ Compare with Wedderburn, [1], page 125. The numbers in square brackets refer to the bibliography.

$\mathfrak{R}(x)$ has dimensionality two less than \mathfrak{M} . For it is defined by the complex equation

$$\det (X - x) = 0.$$

II. \mathfrak{N} , the set of matrices whose characteristic equation has distinct and simple roots.

III. \mathfrak{D} , the set of matrices whose reduced characteristic equation is of lower degree than the characteristic equation.

IV. \mathfrak{S} , the complement of \mathfrak{D} in $\mathfrak{M} - \mathfrak{N}$.

The set $\mathfrak{D} + \mathfrak{S}$ is also $(2n^2 - 2)$ -dimensional. For the coördinates of the matrices which belong to it satisfy the complex equation obtained by setting the discriminant of the characteristic equation equal to zero. These equations define an algebraic locus, so that the following topological theorem is valid.

THEOREM 1.1. *Any matrix of \mathfrak{S} or \mathfrak{D} can be approached by matrices which belong to \mathfrak{N} .*

2. Corresponding to each elementary divisor of X is a pair of matrices called partial idem-potent and nil-potent elements of X . If X does not belong to \mathfrak{D} , these matrices are uniquely defined. If X belongs to \mathfrak{D} , they can be found, but not uniquely. In case they are unique, they are called principal idem-potent and nil-potent elements of X .²

If X belongs to \mathfrak{N} and has the characteristic equation

$$g(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n) = 0,$$

then the idem-potent elements of X are the matrices

$$(2.1) \quad P_i(X) = (X - \lambda_1) \cdots (X - \lambda_{i-1})(X - \lambda_{i+1}) \cdots (X - \lambda_n).$$

The nil-potent elements are all zero in this case.

But in any case the partial idem-potent elements, P_i , and the partial nil-potent elements, Q_i , satisfy the equations

$$(2.2) \quad \begin{aligned} P_i P_j &= P_j P_i = \delta_{ij} P_i \\ P_i Q_j &= Q_j P_i = \delta_{ij} Q_i \\ Q_i Q_j &= Q_j Q_i = \delta_{ij} Q_i^2. \end{aligned}$$

In addition, the important identity,

$$(2.3) \quad X = \sum_{i=1}^v (\lambda_i P_i + Q_i)$$

is also true, v being the number of elementary divisors.

²See [1], pages 27-29, 42. Also [2].

Let $f(x)$ be an analytic function of the complex variable x . It is assumed that everything is known about the function $f(x)$, so that the only problem is to extend the function to \mathcal{M} . Assume that the origin is a regular point of $f(x)$ and that it is represented in a neighborhood of the origin, $|x| < c$, by the power series

$$\phi(x) = \sum_{k=0}^{\infty} a_k x^k.$$

The expression

$$(2.4) \quad \phi(X) = \sum_{k=0}^{\infty} a_k X^k$$

defines a mapping of the part of \mathcal{M} for which the right member converges on to a subset of \mathcal{M} . Therefore it is considered as a part of the extended function.

THEOREM 2.1. $\phi(X)$ is an absolutely continuous function of X on $|X| < c$.

The proof of this theorem is exactly the same as that for the corresponding theorem in function theory. Replacing X by its absolute value reduces 2.4 to a power series in $|X|$.

THEOREM 2.2. If $|X| < c$, then it is true that

$$\phi(X) = \sum_{i=1}^p \left[\phi(\lambda_i) P_i + \sum_{\sigma=1}^n \phi^{(\sigma)}(\lambda_i) Q_i^{\sigma} \frac{1}{\sigma!} \right].$$

The first step of the proof is to show that each characteristic root of X is, in absolute value, less than $|X|$. X transforms the unit sphere in a vector space of n dimensions into an ellipsoid in this space. The absolute value which has been chosen is the square root of the sum of the squares of the principal semi-axes of this ellipsoid. Corresponding to each characteristic root, λ_k , there is a vector, v_k , such that $Xv_k = \lambda_k v_k$. Therefore $|\lambda_k|$ is less than the semi-major axis of this ellipsoid, and thus $|\lambda_k| \leq |X|$.

If X is such that $|X| < c' < c$, for any $\epsilon > 0$ there exists an m (which we take greater than n) such that for $|\lambda| < c$

$$(2.5) \quad \left| \phi^{(\sigma)}(\lambda) - \sum_{\rho=0}^m a_{\rho} \rho(\rho-1) \cdots (\rho-\sigma+1) \lambda^{\rho} \right| < \epsilon, \\ [\sigma = (0, 1, \cdots, n)]$$

and

$$(2.6) \quad \left| \phi(X) - \sum_{\rho=0}^m a_{\rho} X^{\rho} \right| < \epsilon.$$

From (2.2) and (2.3) one gets

$$(2.7) \quad \sum_{\rho=0}^m a_{\rho} X^{\rho} = \sum_{\rho=0}^m \sum_{k=1}^p \sum_{\sigma=0}^{\rho} a_{\rho} \frac{\rho!}{(\rho-\sigma)!} \lambda_k^{\rho-\sigma} \frac{1}{\sigma!} Q_k^{\sigma} P_k.$$

But Q_k is nil-potent. Therefore, powers of Q_k higher than n may be omitted. Thus, by combining (2.5) and (2.7), one gets

$$\left| \sum_{\rho=0}^m a_{\rho} X^{\rho} - \sum_{\sigma=0}^n \sum_{k=1}^p \frac{1}{\sigma!} \phi^{(\sigma)}(\lambda_k) P_k Q_k^{\sigma} \right| < (n+1)\epsilon.$$

Combining this with (2.6) leads to

$$\left| \phi(X) - \sum_{k=1}^p \sum_{\sigma=0}^n \frac{1}{\sigma!} \phi^{(\sigma)}(\lambda_k) P_k Q_k^{\sigma} \right| < (n+2)\epsilon.$$

Since ϵ can be taken arbitrarily small, the theorem is proved.

The statement of this theorem is an identity on the region $|X| < c$. The analytic continuation in \mathcal{M} is based on this identity. The region on which $f(X)$ is defined is extended by assuming that a similar relation is valid in \mathcal{M} .

DEFINITION. If X has the partial idem-potent and nil-potent elements P_k and Q_k associated with the roots λ_k , the matrices of the form

$$(2.8) \quad \sum_{k=1}^p \sum_{\sigma=0}^n \frac{1}{\sigma!} f^{(\sigma)}(\lambda_k) P_k Q_k^{\sigma}$$

will be considered as values of the function,—images of X . Moreover, if Y approaches X , any limit of $f(Y)$ is to be admitted as a value of $f(X)$.

The above definition reduces the matrix function to a single function of n independent variables on \mathcal{N} . But the matrix X can be changed in two distinct ways. One can change the characteristic roots, or he can change the idem-potent elements, P_k . The equation (2.8) shows that, on \mathcal{N} , the P_k are not changed in passing from the argument to the value.

THEOREM 2.3. If X belongs to \mathcal{N} , and if each λ_k is a regular point of $f(x)$, all values of $f(X)$ are given by (2.8).

If X belongs to \mathcal{N} , the P_k are given by (2.1). The λ_k are continuous functions of the coördinates, so that the same applies to the P_k . By hypothesis, $f(x)$ is continuous in the neighborhood of each λ_k . Therefore no limiting process can produce limits not of the form of (2.8).

THEOREM 2.4. If W is non-singular, $f(W^{-1}XW) = W^{-1}f(X)W$.

Proof. It is easy to verify that the similarity transformation can be applied to the power series (2.4) and to the matrices of the form (2.8). All values are obtained from these, or by applying limiting processes to such values. The similarity transformation is continuous. Hence the theorem is established.

THEOREM 2.5. *If X has an elementary divisor of degree m associated with λ and if λ is such that the first, second, \dots $(r-1)$ -th derivatives of $f(x)$ vanish for $x=\lambda$, but $f^{(r)}(\lambda) \neq 0$, then $f(X)$ will have s elementary divisors associated with $f(\lambda)$ where s is the smaller of r and m . These elementary divisors will be of degree $[m/r]$ or $[m/r] + 1$.*

The proof of this theorem depends upon the identity (2.8) and the consideration of the rank of powers of the Q_k associated with this elementary divisor in X . The details are omitted. It was stated above that on \mathcal{N} the elementary divisors were changed only in the root associated with them in passing from the argument to the variable. This theorem states that on \mathcal{S} an elementary divisor of the argument may be broken up by passing to the value of the function. Other changes will appear later.

3. This section is devoted to a discussion of the singularities of the extended function. It will appear that the extended function reflects the singularities inherent in the function and that the extension introduces some singularities if the original function is multiple-valued.

If $f(X)$ is single-valued in a neighborhood of $X=A$, but discontinuous at A , A is called a singular point. A is called a pole of the function if the limit of $f(X)$ is not finite no matter how X approaches A .

If $f(X)$ is multiple-valued, the point A will be called a branch point of $f(X)$ if the number of values of the function is different for the point A and for points in every neighborhood of A .

THEOREM 3.1. *The matrices of \mathcal{S} are singular points of $f(X)$ if, and only if, $f(x)$ is multiple-valued. These points are poles of some branches of $f(X)$.*

In view of Theorem 2.4, it is permissible to give the proof in the most convenient coördinate system. Another simplification is accomplished by using matrices with only the essential parts appearing. Since the elementary divisors enter into the function independently, additional elementary divisors may be added later.

Let $Y = \lambda I + J$, where λ is a regular point of $f(x)$ and J is the matrix all of whose elements are zero except those in the diagonal above the main one, which have the value one. Let $X[h]$ denote the matrix

$$X[h] = \begin{vmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda + h & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda + 2h & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & \lambda + (n-2)h & 1 \\ 0 & 0 & 0 & & 0 & \lambda + (n-1)h \end{vmatrix}.$$

The plan is to compute $f(X[h])$, its limit $f(X[0])$, $f(Y)$ and to compare the last two matrices.

Note that $X[h]$ is a matrix in \mathcal{N} . Therefore its principal idem-potent elements are given by (2.1). Making this computation, one finds that the elements of P_k associated with the root $\lambda + (k-1)h$ of $X[h]$ are either 0 or

$$p_{rs}^k[h] = \frac{1}{(s-r)!} \binom{s-r}{s-k} h^{r-s} (-1)^{s-k}, \quad r \leq k \leq s.$$

Putting these into (2.8) one gets

$$f(X[h]) = \sum_{k=1}^n \phi_{jk}[\lambda + (k-1)h] \left\| \frac{1}{(s-r)!} \binom{s-r}{s-k} (-1)^{s-k} h^{r-s} \right\|$$

where ϕ_{jk} denote the various branches of $f(x)$ in the neighborhood of λ . The above can also be written

$$f(X[h]) = \left\| \frac{h^{r-s}}{(s-r)!} \sum_{k=r}^s \binom{s-r}{s-k} (-1)^{s-k} \phi_{jk}[\lambda + (k-1)h] \right\|.$$

In case all the ϕ_{jk} are the same, the terms added represent the $(s-r)$ -th difference of ϕ_j with respect to the increment h . Because the limit of the ratio of the r -th difference of a function to the r -th power of the interval is the r -th derivative, one gets in this case

$$(3.1) \quad f[X(0)] = \left\| \frac{1}{(s-r)!} \phi_j^{(s-r)}(\lambda) \right\|.$$

When $f(x)$ is single-valued, this reduction is possible.

But in case the ϕ_{jk} are not all the same, one of the elements of $f(X[h])$ with $s = r + 1$, namely

$$\frac{1}{h} [\phi_{j_s}(\lambda + rh) - \phi_{j_s}(\lambda + (r-1)h)],$$

has different values of $f(x)$ in the numerator. Hence, in this case, one gets $f(X[0]) = \infty$.

To complete the proof of the theorem, it is necessary to show that similar limits are obtained by using other paths of approach. It was assumed that λ is a regular point of $f(x)$. Therefore, changing the manner in which the roots of $X[h]$ become equal can have no effect on the limits as long as the path lies in \mathcal{N} . The path can also be deformed in this way: let $V[h]$ be a non-singular-matrix valued function of h which is continuous for $0 \leq h < 1$. Then VXV^{-1} is a continuous function of h . In order for this matrix to approach Y it is sufficient (and necessary) that $V[0]Y = YV[0]$. But, by Theorem 2.4, one has $f(VXV^{-1}) = Vf(X)V^{-1}$ for h different from zero.

Also, a similarity transformation is continuous. Therefore, the limit of $f(VXV^{-1}) = V[0]f(X[0])V[0]^{-1}$. In case $f(X[0])$ is finite, equation (3.1) shows that there is a polynomial such that $p(Y) = f(X[0])$. Therefore, in this case $V[0]$ commutes with $f(X[0])$, and the limit exists independent of the path in \mathcal{N} . On the other hand, if $f(X[0])$ is not finite, the similarity transformation cannot change this property. Therefore since any path in \mathcal{N} can be obtained from the original path by a combination of the above distortions, all paths in \mathcal{N} lead to the same limit.

It can be proved by induction that paths in \mathcal{S} lead to values of the type (2.8) for arguments in \mathcal{S} also. The first step is the above proof for two roots becoming equal. The inductive step can be carried out by using an approximation involving matrices in \mathcal{N} . Let Z_m be a sequence of matrices with elementary divisors of degrees less than n . Choose X_m so that $f(X_m)$ approximates $f(Z_m)$ within an amount ϵ/m . Then the limit of $f(X_m)$ is the same as the limit of $f(Z_m)$. Using the above result concerning the limit of $f(X_m)$, one arrives at the theorem.

THEOREM 3.2. *The points of \mathcal{D} are singular points of $f(X)$ if, and only if, $f(x)$ is multiple-valued. The singularity is of this nature: if X approaches a point of \mathcal{D} along some path the limit of $f(X)$ exists but depends upon the path.*

As before, let $X[h]$ be a point in $\mathcal{N} + \mathcal{S}$ and let its limit, $X[0]$, be a point of \mathcal{D} . Let ϕ_j denote various branches of $f(x)$ in the neighborhood of λ_1 . Then a value of $f(X)$ can be written in the form

$$(3.2) \quad f(X[h]) = \sum_{k=1}^r \phi_{j_k}(\lambda_1 + kh)P_k + \sum_{k=r+1}^v f_k(\lambda_k)P_k.$$

Let $X[0]$ have the form

$$X[0] = \lambda_1 \sum_{k=1}^r P_k + \sum_{k=r+1}^v \lambda_k P_k.$$

Then, on applying (2.8), a value of $f(X[0])$ has the form

$$(3.3) \quad \phi_{j_1}(\lambda_1) \sum_{k=1}^r P_k + \sum_{k=r+1}^v f(\lambda_k)P_k.$$

Note that, if the ϕ_{j_k} are not identical, the limit of $f(X[h])$ is not the expression in (3.3). However, if the ϕ_{j_k} are the same (necessarily true for a single-valued function), the limit of $f(X[h])$ is given by (3.3). Thus the points of \mathcal{D} are singular points of the matrix function.

Now let U be a matrix which commutes with $X[0]$ but not with the individual P_k , ($k = 1, 2, \dots, r$). Moreover, U can be chosen so that it will

commute with a linear combination of these P_k only if the coefficients are all the same number. A similarity transformation is continuous. Therefore one gets

$$\lim_{h \rightarrow 0} f(UX[h]U^{-1}) = U \lim_{h \rightarrow 0} f(X[h])U^{-1}.$$

In case the ϕ_{jk} are not all the same, the coefficients of the P_k will not all be the same number, and U cannot commute with the limit of $f(X[h])$. But notice that X and UXU^{-1} approach $X[0]$ along different paths. Therefore the theorem is established.

COROLLARY. *If $f(x)$ is multiple-valued, and if X is a point of \mathcal{D} , the matrices admitted as values of $f(X)$ are the transforms of the values of form (2.8) by the group of non-singular matrices commutative with X .*

THEOREM 3.3. *Unless X belongs to \mathcal{D} , all values of $f(X)$ are of the form (2.8).*

This result is a combination of Theorem 2.3, the proof of Theorem 3.1, and the definition of the sets \mathcal{N} , \mathcal{S} , and \mathcal{D} . The corollary describes the situation otherwise.

The above discussion concerns the singularities which arise from the extension of functions to matrices. The following theorems concern the inherent singularities of the function.

THEOREM 3.4. *If $x = a$ is a singular point of $f(x)$, the points of $\mathcal{R}(a)$ are singular points of $f(X)$.*

This theorem is important because it states that the point singularity of $f(x)$ is exploded into a $(2n^2 - 2)$ -dimensional singularity for $f(X)$. The possibility of carrying the variable "around" a singularity is preserved. The values obtained by a limiting process applied to (3.2) can also be obtained by carrying a value (3.3) around the proper branch locus of $f(X)$, keeping the argument in \mathcal{D} .

THEOREM 3.5. *If X has an elementary divisor of degree r associated with λ , and if $f^{(s)}(\lambda) = \infty$ for some s less than r , then $f(X) = \infty$.*

This theorem is proved by substituting into (2.8), and then applying Theorem 3.3 and the corollary. This theorem can be applied to show why a nil-potent matrix with two rows has no square root. Such a matrix is a singular point of the function.

4. Let $F(X)$ be defined to be any matrix which satisfies the equation

$$(4.1) \quad p(F(X)) = \sum_{k=0}^m a_k [F(X)]^k = X.$$

In this section, the function $F(X)$ defined here will be compared with the corresponding function $f(X)$ defined in Section two. It will appear that $f(X)$ is identical with the primitive solutions of (4.1). The primitive solutions are those solutions of (4.1) which are not solutions of both (4.1) and a lower degree equation. In making this comparison, certain results of Roth [3] and of Franklin [4] concerning the function $F(X)$ will be used.

THEOREM 4.1. *If $f(x)$ is defined by the equation*

$$p(f(x)) = x,$$

then the matrices $f(X)$ defined in Section two are solutions of

$$(4.2) \quad p(f(X)) = X.$$

Proof. The values of the form (2.8) satisfy (4.2). Also the values of the form of the corollary of Theorem 3.2 satisfy (4.2). All values of $f(X)$ are of one of these types or limits of these types. The operations of addition and multiplication are continuous. Therefore, any limit of matrices which satisfy (4.2) will also satisfy (4.2).

In order to prove the converse relationship, it will be convenient to translate some of Franklin's results into the language of this paper. The solutions, $F(X)$, of (4.1) can be put into three classes:

- A. Solutions which are in turn polynomials in X .
- B. Solutions similar to one of form (2.8) by a matrix commutative with X .
- C. All others. These solutions must have elementary divisors whose degrees differ from the degrees of the divisors of X .

Roth showed that all solutions are of Type A unless X belongs to \mathcal{D} . Franklin showed that, in general, solutions of Type B exist whenever X is derogatory. Furthermore, he showed that solutions of Type C exist only if X is a point of \mathcal{D} and if X has a root λ , such that $p'(f(\lambda)) = 0$, associated with several elementary divisors.

The solutions of Type A are given by (2.8). Solutions of Type B are found in the corollary to Theorem 3.2. Hence it remains to show that the solutions of Type C can be obtained by a limiting process applied to values of the form of (2.8) or of the corollary.

The condition $p'(f(\lambda)) = 0$ is equivalent to the condition that $f'(\lambda) = \infty$. Therefore, solutions of Type C can exist only when X has a root which is a branch point of $f(x)$ associated with several elementary divisors.

The branch point of the function $x^{1/m}$ is typical of any branch point of

order $m - 1$. Therefore, the branch point of this function will be investigated instead of the various branch points of the more general function. Let $X[h]$ be the matrix

$$X[h] = \begin{vmatrix} hI_s + J_s & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdots & hI_s + J_s & 0 & \cdots & 0 \\ 0 & \cdots & 0 & hI_r + J_r & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdots & 0 & 0 & \cdots & hI_r + J_r \end{vmatrix}$$

where I_r is the identity matrix of r rows, $s = r + 1$, and J_r is the matrix of r rows with all elements zero except for 1's in the diagonal above the main one. Let $Y[h]$ have the form

$$Y[h] = \begin{vmatrix} wbI_s & I_s & \cdots & 0 & 0 & \cdots & 0 \\ 0 & w^2bI_s & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & w^kbI_s & E_1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & w^{k+1}bI_r & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ E_2 & 0 & \cdots & 0 & 0 & \cdots & w^mbI_r \end{vmatrix}$$

where w is a primitive m -th root of unity, b is an m -th root of h , and E_1 and E_2 are matrices of the form

$$E_1 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{vmatrix}, \quad E_2 = \begin{vmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{vmatrix}.$$

By a direct expansion it is possible to verify that the m -th power of $Y[h]$ is $X[h]$. The blocks along the diagonal come out very simply, and the coefficients of the other blocks are sums of powers of w which are zero. Since $Y[h]$ is a solution of the equation $Y^m = X$, it is of Type B for all values of h different from zero. Moreover, the equation connecting X and Y will be valid in the limit. The rank of $X[0]$ is $n - m$, but the rank of $Y[0]$ is $n - 1$. Therefore $Y[0]$ is an m -th root of Type C. Thus, by changing the equation slightly, a solution of Type C was obtained as a limit of solutions of Type B.

$X[h]$ is a matrix in \mathcal{D} . So, for each h , there are continua of matrices which satisfy the equation. One of them was selected and called $Y[h]$. Note

that $Y[h]$ is not in \mathcal{D} . The limit of $Y[h]$ is in \mathcal{S} . In taking this limit two things vary; both the continuum from which $Y[h]$ was selected and the relative position of $Y[h]$ in the continuum.

THEOREM 4.2. *If λ is a branch point of $f(x)$ of order $m - 1$, $f(\lambda)$ finite, and if X has m elementary divisors associated with λ (k of them of degree $r + 1$ and the rest of degree r), then there exists a value of $f(X)$ in which $f(\lambda)$ is associated with an elementary divisor of degree $mr + k$.*

The above theorem states that some of the solutions of Type C can be obtained as values of the matrix function $f(X)$. But the only solutions so obtained are those which utilize the complete symmetry of all the values of $f(x)$ which merge at the branch point. Hence the above process can lead only to primitive solutions. However, the non-primitive solutions are primitive solutions of a lower degree equation. Therefore, if the above process yields all primitive solutions, it can be used to get all solutions.

Recall that X is a function, namely a polynomial, of $f(X)$. From this view-point, Theorem 2.5 states that all primitive solutions will have the form specified in the hypothesis of Theorem 5.1. Each continuum of values of $f(X[h])$ leads to a continuum of values of Type C for $f(X[0])$. The various primitive solutions of Type C differ only in the values of $f(\lambda)$ associated with the elementary divisors. But one can achieve this same result by properly choosing the continuum from which $Y[h]$ is chosen. Therefore, any solution of Type C is a value of $f(X)$ as defined in Section two.

THEOREM 4.3. *The function $f(X)$ is the same as the function defined as the primitive solutions of (4.1).*

THE UNIVERSITY OF WISCONSIN.

BIBLIOGRAPHY.

1. Wedderburn, *Lectures on Matrices*, American Mathematical Society Publication.
2. Wegner, "Über die Frobeniusschen Kovarianten," *Monatshefte für Mathematik und Physik*, Bd. 40, pp. 201-208.
3. Roth, "A solution of the matrix equation $P(X) = A$," *Transactions of the American Mathematical Society*, vol. 30, pp. 579-596.
4. Franklin, "Algebraic matrix equations," *Journal für Mathematik und Physik*, Bd. 10, pp. 289-314.
5. Cipolla, "Sulle matrici espressioni analitiche di un'altra," *Rendiconti Circolo Matematico de Palermo*, vol. 56, pp. 144-154.

LINEAR DIFFERENTIAL INVARIANCE UNDER AN OPERATOR RELATED TO THE LAPLACE TRANSFORMATION.*¹

By EARL D. RAINVILLE.

1. Introduction. The Laplace integral transformation²

$$(1) \quad \mathcal{L}\{F(t)\} \equiv \int_0^\infty e^{-st} F(t) dt = f(s),$$

is one which associates with each function $F(t)$ of sufficient regularity another function $f(s)$. Elementary known³ properties of the operator \mathcal{L} include

$$(2) \quad \mathcal{L}\left\{\frac{d^n F(t)}{dt^n}\right\} = s^n f(s) - \sum_{k=0}^{n-1} s^{n-1-k} \left(\frac{d^k F}{dt^k}\right)_{t=0},$$

and

$$(3) \quad \mathcal{L}\{t^k F(t)\} = (-1)^k \frac{d^k}{ds^k} f(s).$$

The Laplace transformation has important applications⁴ to the solution of boundary value problems in ordinary and partial linear differential equations. The operator \mathcal{L} often transforms one differential equation into another which is more readily solved, one which, indeed, may even be algebraic. The transformed equation may be of higher order, or otherwise more complicated, than the original. Finally, we see that many equations do not change form in any essential way when subjected to the operator \mathcal{L} . One such equation is

$$(4) \quad \frac{d^2 F}{dt^2} + t^2 F = 0,$$

* Received August 15, 1939.

¹ Presented to the Society Nov. 26, 1938 under a slightly different title.

² For an extensive treatment of this transformation see G. Doetsch, *Theorie und Anwendung der Laplace-Transformation*, Berlin, 1937.

³ Enzo Levi has shown that the case $n=1$ of equation (2) above, together with certain conditions on $F(t)$ and its transform is sufficient to characterize the operator completely. For the precise result see his paper, "Proprietà caratteristiche della trasformazione di Laplace," *Rend. Accad. Lincei*, (6), vol. 24 (1936), pp. 422-426.

⁴ See, for example, R. V. Churchill, "The solution of linear boundary value problems in physics by means of the Laplace transformation": I, *Mathematische Annalen*, vol. 114 (1937), pp. 591-613; II, *Mathematische Annalen*, vol. 115 (1938), pp. 720-739. See also his paper, "On the problem of temperatures in a non-homogeneous bar with discontinuous initial temperatures," *American Journal of Mathematics*, vol. 61 (1939), pp. 651-664, in which the Laplace transformation is used to establish a uniqueness theorem.

for which the transformed equation is

$$(5) \quad \frac{d^2 f}{ds^2} + s^2 f = s(F)_{t=0} + \left(\frac{dF}{dt} \right)_{t=0},$$

as may be seen from (2) and (3).

Our fundamental problem is suggested by the fact that (4) is essentially invariant under \mathcal{L} . In order to use only those properties of \mathcal{L} which are concerned in that invariance we introduce another operator σ and study σ instead of \mathcal{L} .

DEFINITION 1. Let $D \equiv d/dx$ be the usual symbol for differentiation with respect to x ; let $D^0 \equiv 1$. Then any polynomial in D and x will be called a *linear differential operator of type P*.

DEFINITION 2. Let k, n, k_s, n_s ; $s = 1, 2, \dots$, be non-negative integers. We define ${}^s\sigma$ as a linear operator on linear differential operators of type P by

$$(6) \quad \sigma x^k D^n = (-1)^k D^k x^n,$$

and

$$(7) \quad \sigma \left[\sum_s a_s x^{k_s} D^{n_s} \right] = \sum_s a_s \sigma (x^{k_s} D^{n_s}),$$

where the a_s are any constants.

It should be noted that by $D^k x^n$ we mean that differential operator which, acting upon a function F , yields the k -th derivative of the product $x^n F$.

It is of some value to keep in mind one aspect of the nature of σ . Consider a given function $f(x)$, taken to be single valued for the present purpose. We may associate with $f(x)$ an operator f which transforms each number x of a certain set of numbers into another number $f(x)$ in another, or the same, set of numbers. We call f an operator of class one. Next consider D . The operator D transforms each function of a certain set of functions into another function of x . Further, if D operate on numbers, the result is trivial; i.e., D transforms every number into the same number, zero. Hence, we call D an operator of class two, noting that in a sense D must operate on operators of class one to give non-trivial results. Now consider σ . This operator is defined above in such a way that it transforms each linear differential operator into another linear differential operator. Essentially σ needs to operate on operators of class two to give non-trivial results. We call σ an operator of class three.

An adjoint operator may be defined such that it changes a linear dif-

* Essentially this definition is to be found in S. Pincherle and U. Amaldi, *Operazioni distributive*, Bologna, 1901, p. 361.

ferential operator, not necessarily of type P , into its adjoint linear differential operator. This adjoint operator⁶ is of class three in the above sense.

2. Results. Some useful, not all new, properties of σ are obtained. Two linear bases are found for the set of linear differential operators invariant under σ . Two invariant second order differential operators are found to form a fundamental system of invariant operators; i. e., any invariant operator may be expressed as a polynomial in these two operators. A linear basis is exhibited for what are called t -variants (Definition 3) with respect to σ .

Linear operational equations in σ are completely solved in the case of constant coefficients. This is done with the aid of two theorems on the representation of linear differential operators in terms of t -variants or of invariants and pseudo-invariants. The same tools are useful in the solution of linear operational equations in σ with variable (linear differential operational) coefficients, as is demonstrated in the example worked out in Section 10.

In Section 9 certain results are specialized to yield a classification of the differential equations, such as (4) above, invariant under σ .

3. Preliminary definitions and lemmas. Since the only linear differential operators to enter this study are of type P , we shall often hereafter omit mention of this restriction.

DEFINITION 3. If y is a linear differential operator such that $\sigma y = ty$ where $t^4 = 1$, then y will be called a *linear differential t -variant* with respect to σ . A 1-variant will be referred to on occasion as an *invariant* and a (-1) -variant may be called a *pseudo-invariant*.

DEFINITION 4. The *degree* and the *order* of a linear differential operator are respectively the highest power of the independent variable and the order of the highest ordered derivative appearing explicitly in the operator.

LEMMA 1. If y is a linear differential operator, then the degree of $\sigma y =$ the order of y and the order of $\sigma y =$ the degree of y .

This lemma is an immediate consequence of the definition of σ . The application of Lemma 1 leads to

⁶E. D. Rainville, "Adjoints of linear differential operators," *American Mathematical Monthly*, vol. 46 (1939), pp. 623-627. For relations between σ and the adjoint operator, see L. Schlesinger, *Handbuch der Linearen Differentialgleichungen*, Leipzig, 1895, vol. 1, p. 426 and E. D. Rainville, "A discrete group arising in the study of differential operators," as yet unpublished.

LEMMA 2. For a linear differential operator to be t -variant with respect to σ it is necessary that its degree equal its order.

LEMMA 3. The linear differential operators $A_1 = D^2 + x^2$ and $A_2 = x^2 D^2 + 2xD$, are invariant with respect to σ .

We prove Lemma 3 by direct evaluation of σA_1 and σA_2 .

$$\sigma A_1 = x^2 + D^2 = A_1.$$

$$\sigma A_2 = D^2 x^2 - 2Dx = x^2 D^2 + 4xD + 2 - 2xD - 2 = A_2.$$

THEOREM 1. If A and B are linear differential operators, then

$$\sigma(AB) = (\sigma A)(\sigma B).$$

Let $v = x^k D^n$, then

$$\sigma(xv) = \sigma(x^{k+1} D^n) = (-1)^{k+1} D^{k+1} x^n = -D\sigma(x^k D^n),$$

so that we have

$$(8) \quad \sigma(xv) = (\sigma x)(\sigma v).$$

Next,

$$\begin{aligned} \sigma(Dv) &= \sigma(x^k D^{n+1} + kx^{k-1} D^n) = (-1)^k D^k x^{n+1} + (-1)^{k-1} k D^{k-1} x^n \\ &= (-1)^k x D^k x^n + (-1)^k k D^{k-1} x^n + (-1)^{k-1} k D^{k-1} x^n = (-1)^k x D^k x^n = x(\sigma v). \end{aligned}$$

Then

$$(9) \quad \sigma(Dv) = (\sigma D)(\sigma v).$$

Theorem 1 follows directly from (8) and (9). Further,

$$(10) \quad \sigma^2(AB) = \sigma[(\sigma A)(\sigma B)] = (\sigma^2 A)(\sigma^2 B),$$

and, for any integral $k \geq 0$, $\sigma^k(AB) = (\sigma^k A)(\sigma^k B)$.

LEMMA 4. If $v = x^k D^n$, then $\sigma^2 v = (-1)^{k+n} v$.

By (10) above

$$\sigma^2 v = (\sigma^2 x^k)(\sigma^2 D^n) = [\sigma(-1)^k D^k][\sigma x^n] = (-1)^k x^k (-1)^n D^n = (-1)^{k+n} v.$$

Lemma 4 itself leads at once to ⁷

THEOREM 2. If y is a linear differential operator, then $\sigma^4 y = y$.

4. First classification of invariants. Direct application of Theorem 1 and Lemma 3 yields

THEOREM 3. Any linear combination of terms of the type

$$A_1^{m_1} A_2^{m_2} A_1^{m_3} \cdots A_1^{m_{s-1}} A_2^{m_s}$$

⁷ Theorem 2 appears in Pincherle and Amaldi, *loc. cit.*, p. 357.

in which $m_i; i = 1, 2, \dots, s$, are non-negative integers, is a linear differential invariant with respect to σ .

Next we obtain a simple necessary condition for invariance under σ .

LEMMA 5. *A necessary and sufficient condition that a linear differential operator be invariant under σ^2 is that, for each term $a_{kn}x^kD^n$ of the operator, $k \equiv n \pmod{2}$.*

This follows at once from Lemma 4. Noting that invariance under σ implies invariance under σ^2 , we have a necessary condition for the former.

THEOREM 4. *For each term $a_{kn}x^kD^n$ of a linear differential invariant with respect to σ it is true that $k \equiv n \pmod{2}$.*

DEFINITION 5. The leading term of a linear differential operator is the non-vanishing term of highest degree among those terms of highest order in the operator.

It will prove useful to note that, since the order of the operator is the order of its leading term, we have

LEMMA 6. *In a linear differential t -variant the degree of the leading term does not exceed its order.*

DEFINITION 6. By linear differential invariants of type H we mean the set of invariant operators

$$(11) \quad A_1^{n-k}A_2^k; \quad 0 \leq k \leq n,$$

and

$$(12) \quad \frac{1}{4(n-k)(k+1)} [A_1^{n-k}A_2^{k+1} - A_2^{k+1}A_1^{n-k}]; \quad 0 \leq k < n.$$

LEMMA 7. *The leading term of $A_1^{n-k}A_2^k; 0 \leq k \leq n$, is $x^{2k}D^{2n}$.*

This follows at once from the definitions of A_1 and A_2 .

LEMMA 8. *The leading term of*

$$\frac{1}{4(n-k)(k+1)} [A_1^{n-k}A_2^{k+1} - A_2^{k+1}A_1^{n-k}]; \quad 0 \leq k < n,$$

is $x^{2k+1}D^{2n+1}$.

In the proof of Lemma 8 we shall use the convention that

$$y = a_{\delta\epsilon}x^\delta D^\epsilon + a_{\delta'\epsilon'}x^{\delta'}D^{\epsilon'} + \dots$$

means that y is a linear differential operator with leading term $a_{\delta\epsilon}x^\delta D^\epsilon$ and that the leading term of $(y - a_{\delta\epsilon}x^\delta D^\epsilon)$ is $a_{\delta'\epsilon'}x^{\delta'}D^{\epsilon'}$.

With the above convention note that the formula

$$(13) \quad A_2^k = x^{2k} D^{2k} + 2k^2 x^{2k-1} D^{2k-1} + \dots$$

holds for $k=1$ and in a trivial sense for $k=0$. Assume (13) to hold for some k . Then

$$\begin{aligned} A_2^{k+1} &= x^{2k+2} D^{2k+2} + (4k + 2k^2 + 2) x^{2k+1} D^{2k+1} + \dots \\ &= x^{2k+2} D^{2k+2} + 2(k+1)^2 x^{2k+1} D^{2k+1} + \dots, \end{aligned}$$

and it follows by induction that (13) is true for any $k \geq 0$. Now

$$(14) \quad A_1^{n-k} A_2^k = x^{2k} D^{2n} + 2k(2n-k) x^{2k-1} D^{2n-1} + \dots$$

holds for $n=k \geq 0$. Assume (14) to hold for some pair of numbers n, k . Then

$$\begin{aligned} (15) \quad A_1^{n-k+1} A_2^k &= x^{2k} D^{2n+2} + [4k + 2k(2n-k)] x^{2k-1} D^{2n+1} + \dots \\ &= x^{2k} D^{2n+2} + 2k[2(n+1) - k] x^{2k-1} D^{2n+1} + \dots, \end{aligned}$$

so that by induction (14) holds for any $n \geq k \geq 0$.

In view of (13) the application of A_2^k to A_1^{n-k+1} is seen to yield

$$(16) \quad A_2^k A_1^{n-k+1} = x^{2k} D^{2n+2} + 2k^2 x^{2k-1} D^{2n+1} + \dots,$$

for $n \geq k \geq 0$. Combining (15) and (16) with k replaced by $(k+1)$, we have as the leading term of $[A_1^{n-k} A_2^{k+1} - A_2^{k+1} A_1^{n-k}]$; $0 \leq k < n$, the expression $4(k+1)(n-k) x^{2k+1} D^{2n+1}$, so that Lemma 8 is established.

THEOREM 5. *A necessary and sufficient condition that there exist a linear differential invariant with respect to σ with leading term $x^\delta D^\epsilon$ is that either*

$$\delta \equiv \epsilon \equiv 0 \pmod{2}, \quad 0 \leq \delta \leq \epsilon,$$

or

$$\delta \equiv \epsilon \equiv 1 \pmod{2}, \quad 1 \leq \delta < \epsilon.$$

Lemmas 7 and 8 exhibit linear differential invariants for each leading term indicated in Theorem 5. We proceed to show that no linear differential invariant can exist with leading term not proportional to one of those indicated in Theorem 5. By Theorem 4 we must have $\delta \equiv \epsilon \pmod{2}$. By Lemma 6 we must have $\delta \leq \epsilon$. We have left the one case $\delta = \epsilon = 2h + 1$ and we consider that now. If a linear differential operator y had for its leading term $x^{2h+1} D^{2h+1}$, then σy would have for its leading term $(-x^{2h+1} D^{2h+1})$, and y could not be invariant. This concludes the proof of Theorem 5. We proceed to the main result of this section.

THEOREM 6. *Any linear differential invariant with respect to σ is a linear combination of linear differential invariants of type H.*

Stated in another way, we show that the invariants of type H form a linear (infinite) basis for the algebra whose elements are the invariants with respect to σ .

Let y be any linear differential invariant with leading term $a_{\delta\epsilon}x^\delta D^\epsilon$, where, of course, δ and ϵ are subject to the restrictions of Theorem 5. By Lemmas 7 and 8 there exists a linear differential invariant of type H with leading term $x^\delta D^\epsilon$. Hence we see that there exists a linear combination of y and a linear differential invariant of type H (with coefficient of y not zero) which is invariant under σ and is such that its leading term is either (a) of lower order than the leading term of y , or (b) of the same order and of lower degree than the leading term of y . Repetition of this argument shows that there exists an identically vanishing linear combination (with coefficient of y not zero) of y and linear differential invariants of type H . Thus Theorem 6 is established.

Next we note that, since no two of the linear differential invariants of type H have proportional leading terms, it follows that the linear differential invariants of type H are linearly independent.

The preceding work, particularly Theorem 6, shows that A_1 and A_2 form a fundamental system of invariant differential operators in the sense of

THEOREM 7. *Any linear differential operator invariant with respect to σ may be expressed as a polynomial in A_1 and A_2 .*

Of course, Theorem 3 has already stated that any polynomial in A_1 and A_2 is invariant under σ . Since A_1 is not commutative with A_2 , the word polynomial is used here in the sense of linear combinations of operators of the type exhibited in Theorem 3.

5. Second classification of invariants.

LEMMA 9. *A necessary and sufficient condition for*

$$I_{kn} = x^k D^n + \sigma(x^k D^n); \quad 0 \leq k, n,$$

to be a linear differential invariant with respect to σ is that $k \equiv n \pmod{2}$.

Noting that

$$\sigma I_{kn} = \sigma(x^k D^n) + \sigma^2(x^k D^n),$$

and recalling Lemma 4 we see that Lemma 9 follows at once.

DEFINITION 7. By *linear differential invariants of type J* we mean the set of invariant operators

$$\begin{aligned} \text{and} \quad I_{kn}; \quad k \equiv n \equiv 0 \pmod{2}, \quad 0 \leq k \leq n, \\ I_{kn}; \quad k \equiv n \equiv 1 \pmod{2}, \quad 1 \leq k < n. \end{aligned}$$

Note that in the set of linear differential invariants of type J whenever $k \neq n$ the leading term of I_{kn} is $x^k D^n$; if $k = n$, then k and n are even and the leading term of I_{nn} is $2x^n D^n$. Hence it is evident that the linear differential invariants of type J are linearly independent.

Since all leading terms permitted by Theorem 5 are included in type J , we may follow the line of reasoning used to prove Theorem 6 and thus demonstrate

THEOREM 8. *Any linear differential operator invariant with respect to σ may be expressed linearly in terms of linear differential invariants of type J .*

See also the remark directly below Theorem 6.

6. A classification of t -variants. We shall briefly indicate a classification of t -variants similar to the above second classification of invariants. This done, we may consider t -variants completely specified and may proceed to two representation theorems with the aid of which we solve linear operational equations in σ .

From Lemma 4 of Section 3 we get

LEMMA 10. *A necessary and sufficient condition that a linear differential operator be pseudo-invariant with respect to σ^2 is that, for each term $a_{kn}x^k D^n$ of the operator, $k \equiv n + 1 \pmod{2}$.*

Let $i = \sqrt{-1}$. If $t = i$ or if $t = i^3$, then $t^2 = -1$ and any corresponding linear differential t -variant with respect to σ is pseudo-invariant with respect to σ^2 . If $t = -1$, then we have actual invariance with respect to σ^2 . Hence Lemmas 5 and 10 lead to

THEOREM 9. *For each term $a_{kn}x^k D^n$ of a linear differential t -variant with respect to σ it is true that $k \equiv n + \frac{1}{2}(1 - t^2) \pmod{2}$.*

LEMMA 11. *A necessary and sufficient condition that*

$$I_{kn}^{(t)} = x^k D^n + t^3 \sigma(x^k D^n), \quad t^4 = 1,$$

be a t -variant with respect to σ is that $k \equiv n + \frac{1}{2}(1 - t^2) \pmod{2}$.

Since

$$\sigma I_{kn}^{(t)} = \sigma(x^k D^n) + t^3 \sigma^2(x^k D^n)$$

and

$$t I_{kn}^{(t)} = t x^k D^n + \sigma(x^k D^n),$$

a necessary and sufficient condition for the equality of $\sigma I_{kn}^{(t)}$ and $t I_{kn}^{(t)}$ is that

$\sigma^2(x^k D^n) = t^2 x^k D^n$. By Lemma 4 this is equivalent to $t^2 = (-1)^{k+n}$ or to $k \equiv n + \frac{1}{2}(1 - t^2) \pmod{2}$.

Considerations similar to those used in the proof of Theorem 5 yield a proof (omitted here) of

THEOREM 10. *A necessary and sufficient condition that there exist a linear differential t -variant with respect to σ with leading term $x^\delta D^\epsilon$ is that either*

$$\begin{aligned} \delta &\equiv \epsilon + \frac{1}{2}(1 - t^2) \equiv 0 \pmod{2}, & 0 \leq \delta < \epsilon + \frac{1}{4}(1 + t^2)(1 + t), \\ \text{or} & & \\ \delta &\equiv \epsilon + \frac{1}{2}(1 - t^2) \equiv 1 \pmod{2}, & 1 \leq \delta < \epsilon + \frac{1}{4}(1 + t^2)(1 - t). \end{aligned}$$

DEFINITION 8. By *linear differential t -variants of type J* we mean the set of t -variant operators

$$\begin{aligned} I_{kn}^{(t)}; & \quad k \equiv n + \frac{1}{2}(1 - t^2) \equiv 0 \pmod{2}, \quad 0 \leq k < n + \frac{1}{4}(1 + t^2)(1 + t), \\ \text{and} & \\ I_{kn}^{(t)}; & \quad k \equiv n + \frac{1}{2}(1 - t^2) \equiv 1 \pmod{2}, \quad 1 \leq k < n + \frac{1}{4}(1 + t^2)(1 - t). \end{aligned}$$

It can be seen that the linear differential t -variants of type J are linearly independent. Reasoning parallel to that used to prove Theorem 6 will demonstrate

THEOREM 11. *Any linear differential t -variant with respect to σ may be expressed linearly in terms of linear differential t -variants of type J .*

See also the remark directly below Theorem 6.

7. Representation theorems. We shall prove the following two theorems on the representation of linear differential operators of type P .

THEOREM 12. *Any linear differential operator of type P may be represented in one, and only one, way in the form*

$$(17) \quad I + P + Q + W$$

where I is an invariant, P a pseudo-invariant, Q an i -variant, and W an i^3 -variant.

THEOREM 13. *Any linear differential operator of type P may be represented in one, and only one, way in the form*

$$(18) \quad I_1 + P_1 + x(I_2 + P_2) + D(I_3 + P_3)$$

where I_1, I_2, I_3 are invariants and P_1, P_2, P_3 are pseudo-invariants.

In order to picture more clearly the relation between Theorems 12 and 13, let us consider the operator $2x^2D$. We may write

$$2x^2D = (-ixD^2 + x^2D - 2iD) + (ixD^2 + x^2D + 2iD),$$

where $Q = -ixD^2 + x^2D - 2iD$ and $W = ixD^2 + x^2D + 2iD$ are respectively an i -variant and an i^3 -variant. Here, though the operator $2x^2D$ is real, the representation (17) introduces the imaginary unit. We may, on the other hand, write

$$2x^2D = x[(-1) + (2xD + 1)],$$

where $I_2 = -1$ and $P_2 = 2xD + 1$ are respectively an invariant and a pseudo-invariant. Hence, using (18) the representation of $2x^2D$ is "real." The representations (17) and (18) play roles corresponding to the two solutions $F = a_1e^{ix} + a_2e^{-ix}$ and $F = c_1 \cos x + c_2 \sin x$ of the differential equation $(D^2 + 1)F = 0$.

The example in Section 10 illustrates the fact that (18) may on occasion have considerable advantage over the apparently simpler and more natural representation (17).

The representation (17) is essentially a result of the fact that σ satisfies the operational equation $\sigma^4 = E$, the identity.

Proof of Theorem 12. First we give an explicit expression for any term $x^k D^n$ of a linear differential operator in the manner desired. Let k, n be non-negative integers. Then, using the notation of Lemma 11, we have

$$(19) \quad x^k D^n = \frac{1}{4}[1 + (-1)^{k+n}][I_{kn}^{(1)} + I_{kn}^{(-1)}] + \frac{1}{4}[1 - (-1)^{k+n}][I_{kn}^{(i)} + I_{kn}^{(-i)}].$$

Because of the linearity of the operators uniqueness of (17) will follow if we show that

$$(20) \quad I + P + Q + W = 0,$$

with the notation of Theorem 12, implies $I = P = Q = W = 0$.

If we operate on (20) with σ^2 we find

$$(21) \quad I + P - Q - W = 0.$$

From (20) and (21) we have $I + P = 0$. Operating on this with σ , we get $I - P = 0$. Hence $I = P = 0$. But (20) and (21) also lead to $Q + W = 0$, from which it follows that $iQ - iW = 0$. Hence $Q = W = 0$ and the proof of Theorem 12 is complete.

Proof of Theorem 13. In order to show the existence of the representation (18) we write the identity, for k and n non-negative integers,

$$(22) \quad \begin{aligned} x^k D^n &= \frac{1}{4} [1 + (-1)^{k+n}] [I_{kn}^{(1)} + I_{kn}^{(-1)}] \\ &+ \frac{1}{4} (\operatorname{sgn} k) [1 - (-1)^{k+n}] x [I_{k-1,n}^{(1)} + I_{k-1,n}^{(-1)}] \\ &+ \frac{1}{4} (1 - \operatorname{sgn} k) [1 - (-1)^n] D [I_{0,n-1}^{(1)} + I_{0,n-1}^{(-1)}], \end{aligned}$$

in which $\operatorname{sgn} k$ is the usual signum function with argument k .

In order to prove the uniqueness of the representation (18), we may follow in part the method used in the proof of Theorem 12. Assume

$$(23) \quad I_1 + P_1 + x(I_2 + P_2) + D(I_3 + P_3) = 0,$$

with the notation as in Theorem 13. Operating on (23) with σ^2 we find

$$(24) \quad I_1 + P_1 - x(I_2 + P_2) - D(I_3 + P_3) = 0.$$

From these two equations $I_1 + P_1 = 0$ and hence $I_1 = P_1 = 0$ follow. We are left with

$$(25) \quad x(I_2 + P_2) + D(I_3 + P_3) = 0.$$

Here the method of proof digresses from that above. We recall that the degree of a linear differential invariant equals its order and that the same is true of a pseudo-invariant (Lemma 2). Hence the degree of $(I_2 + P_2)$ equals its order, say n_2 . Also the degree of $(I_3 + P_3)$ equals its order, say n_3 . Since the degrees and the orders of the two terms of (25) must be respectively equal, we have $n_2 + 1 = n_3$ and $n_2 = n_3 + 1$. But no pair of values n_2 and n_3 can satisfy these relations. Thus we have $x(I_2 + P_2) = 0$ and $D(I_3 + P_3) = 0$. It follows at once that $I_2 = P_2 = I_3 = P_3 = 0$.

Suppose an operator y is represented in each of the forms (17) and (18). It is then easy to obtain I , P , Q , and W in terms of I_1 , I_2 , I_3 , P_1 , P_2 , P_3 . We find $I = I_1$, $P = P_1$,

$$(26) \quad Q = \frac{1}{2} (D - ix) (I_3 + iI_2) + \frac{1}{2} (D + ix) (P_3 - iP_2),$$

and

$$(27) \quad W = \frac{1}{2} (D + ix) (I_3 - iI_2) + \frac{1}{2} (D - ix) (P_3 + iP_2).$$

An examination of (20) and (22) leads us to believe that the determination of I_2 , I_3 , P_2 , P_3 in terms of Q and W is a term by term affair not to be written in such simple forms as (26) and (27).

8. Linear operational equations in σ : constant coefficients. Next we consider linear operational equations in σ in analogy to linear differential equations in D . The unknown to be determined is now a linear differential operator. Since $\sigma^4 y = y$ for y any linear differential operator, we consider only operational equations of "order" ≤ 3 in σ .

First we treat the homogeneous equation

$$(28) \quad a_3 \sigma^3 y + a_2 \sigma^2 y + a_1 \sigma y + a_0 y = 0,$$

where a_3, a_2, a_1, a_0 are constants. We may operate on (28) with σ and obtain

$$a_2 \sigma^3 y + a_1 \sigma^2 y + a_0 \sigma y + a_3 y = 0,$$

$$a_1 \sigma^3 y + a_0 \sigma^2 y + a_3 \sigma y + a_2 y = 0,$$

$$a_0 \sigma^3 y + a_3 \sigma^2 y + a_2 \sigma y + a_1 y = 0.$$

For (28) to have a solution other than $y = 0$ it is necessary that the determinant of the coefficients of $y, \sigma y, \sigma^2 y, \sigma^3 y$, in the above equations vanish. This leads at once to

THEOREM 14. *A necessary condition that there exist a non-vanishing linear differential operator y which satisfies*

$$(28) \quad a_3 \sigma^3 y + a_2 \sigma^2 y + a_1 \sigma y + a_0 y = 0,$$

where a_3, a_2, a_1, a_0 are constants, is that

$$(29) \quad (a_3 + a_2 + a_1 + a_0)(a_3 - a_2 + a_1 - a_0)[(a_3 - a_1)^2 + (a_2 - a_0)^2] = 0.$$

If we now put $y = I + P + Q + W$, Equation (28) yields

$$(30) \quad \begin{aligned} (a_3 + a_2 + a_1 + a_0)I &= \Delta_1 I = 0, \\ (a_3 - a_2 + a_1 - a_0)P &= \Delta_2 P = 0, \\ [(a_3 - a_1) - i(a_2 - a_0)]Q &= \Delta_3 Q = 0, \\ [(a_3 - a_1) + i(a_2 - a_0)]W &= \Delta_4 W = 0, \end{aligned}$$

with the aid of Theorem 12. By means of Equations (30) we are able to determine the general solution of (28). For example, if $\Delta_1 = \Delta_3 = \Delta_4 = 0$ and $\Delta_2 \neq 0$, then the general solution of (28) is $y = I + Q + W$, where I is any invariant, Q any i -variant and W is any i^3 -variant.

We turn now to the non-homogeneous case.

THEOREM 15. *Let a_3, a_2, a_1, a_0 be constants such that $\Delta_1 \Delta_2 \Delta_3 \Delta_4 \neq 0$, where the Δ 's are as defined in (30). Let I, P, Q, W be as defined in Theorem*

12. Then there exists one, and only one, linear differential operator which satisfies the equation

$$(31) \quad a_3\sigma^3y + a_2\sigma^2y + a_1\sigma y + a_0y = I + P + Q + W,$$

namely

$$(32) \quad y = \frac{I}{\Delta_1} - \frac{P}{\Delta_2} + i \frac{Q}{\Delta_3} - i \frac{W}{\Delta_4}.$$

That (32) is a solution of (31) may be seen by direct substitution. If there were two distinct solutions of (31), the non-vanishing difference of those solutions would satisfy (28), the homogeneous equation. In view of the inequality of Theorem 15 and the necessary condition in Theorem 14, this is impossible.

Equation (32) is readily altered to fit the case where the homogeneous equation also has a solution. Let us suppose, for example, that in (31) we find $\Delta_2 = \Delta_4 = 0$, $\Delta_1\Delta_3 \neq 0$. Then there exists no solution of (31) unless $P = 0$ and $W = 0$. If these conditions are satisfied, the general solution of (31) is

$$y = \frac{I}{\Delta_1} + P_1 + i \frac{Q}{\Delta_3} + W_1,$$

where P_1 is any pseudo-invariant, W_1 any i^3 -variant and the other symbols are as in Theorem 15. There is here a noticeable resemblance to the general solution of a non-homogeneous ordinary linear differential equation.

If in Theorem 15 we use the representation of Theorem 13, instead of that of Theorem 12, we need only to replace (31) by

$$(31') \quad a_3\sigma^3y + a_2\sigma^2y + a_1\sigma y + a_0y = I_1 + P_1 + x(I_2 + P_2) + D(I_3 + P_3),$$

and (32) by

$$(32') \quad y = \frac{I_1}{\Delta_1} - \frac{P_1}{\Delta_2} - \frac{a_3 - a_1}{\Delta_3\Delta_4} [D(I_2 - P_2) - x(I_3 - P_3)] \\ - \frac{a_2 - a_0}{\Delta_3\Delta_4} [x(I_2 + P_2) + D(I_3 + P_3)].$$

9. Linear differential equations invariant under σ . We have incidentally solved the problem of determining what linear differential equations are invariant under σ . Theorem 14 and the remarks following it yield at once

THEOREM 16. Let y be a linear differential operator of type P and let F be an undetermined function of x . Then a necessary and sufficient condition that $yF = 0$ be invariant under σ is that y be a t -variant with respect to σ .

This is an interpretation of the fact that for $\sigma y = cy$ to have a solution for constant c it is necessary and sufficient that $c^4 = 1$.

By means of the representation (18) of Theorem 13 we are able to state this result in another form sometimes more useful.

THEOREM 17. *Let y be a linear differential operator of type P and let F be an undetermined function of x . Then a necessary and sufficient condition that $yF = 0$ be invariant under σ is that y be expressible in one of the four forms (33)–(36);*

$$(33) \quad y = I_1,$$

$$(34) \quad y = P_1,$$

$$(35) \quad y = x(I_2 + P_2) + iD(I_2 - P_2),$$

$$(36) \quad y = x(I_2 + P_2) - iD(I_2 - P_2),$$

where I_1, I_2 are invariants, P_1, P_2 are pseudo-invariants, and $i = \sqrt{-1}$.

10. Linear operational equations in σ : variable coefficients. We may generalize equation (31) to the case where the coefficients are themselves linear differential operators. We consider

$$(37) \quad a_3(\sigma^3 y)b_3 + a_2(\sigma^2 y)b_2 + a_1(\sigma y)b_1 + a_0 y b_0 = A,$$

where the a_i, b_i ; $i = 0, 1, 2, 3$, and A are known linear differential operators and y is to be determined. We shall illustrate our two methods of attack on (37) by means of a numerical example. Consider

$$(38) \quad x\sigma y - Dy = 0.$$

If we use the representation $y = I + P + Q + W$ as indicated in Theorem 12, we find that

$$(39) \quad x(I - P) + ix(Q - W) - D(I + P) - D(Q + W) = 0.$$

Operating on (39) with $\sigma, \sigma^2, \sigma^3$, and combining the resulting equations, we readily obtain $I = 0, P = 0$, and

$$(40) \quad (D - ix)Q + (D + ix)W = 0.$$

Further, if we substitute $y = Q + W$ into (38) we get (40). Hence, the general solution of (38) is $y = Q + W$ where Q and W are respectively any i -variant and any i^3 -variant subject to the restriction (40).

Let us now attack (38) with the representation made available by Theorem 13. In this case the solution appears in a more satisfactory form. Let $y = I_1 + P_1 + x(I_2 + P_2) + D(I_3 + P_3)$. Then from (38) we get

$$(41) \quad (x-D)I_1 - (x+D)P_1 \\ - (2xD+1)I_2 - P_2 + x^2(I_3 - P_3) - D^2(I_3 + P_3) = 0.$$

Using σ on (41) we arrive at

$$(42) \quad -(x+D)I_1 - (D-x)P_1 \\ + (2xD+1)I_2 + P_2 + D^2(I_3 + P_3) - x^2(I_3 - P_3) = 0.$$

Equations (41) and (42) combine to yield $2D(I_1 + P_1) = 0$, hence $I_1 = P_1 = 0$. We return to (41) which has become

$$(43) \quad (2xD+1)I_2 + P_2 - x^2(I_3 - P_3) + D^2(I_3 + P_3) = 0.$$

Substituting for P_2 from (43) into the assumed expression for y we get

$$y = -2x^2DI_2 - (xD^2 - D - x^3)I_3 - (xD^2 - D + x^3)P_3.$$

Since I_2 and I_3 may be any invariants and P_3 any pseudo-invariant, we shall write

$$(44) \quad y = x^2DI_4 + (xD^2 - D - x^3)I_5 + (xD^2 - D + x^3)P_5.$$

Then

$$x\sigma y = x(D^2x)I_4 + x(-Dx^2 - x + D^3)I_5 - x(-Dx^2 - x - D^3)P_5 \\ = x(xD^2 + 2D)I_4 + x(D^3 - x^2D - 3x)I_5 + x(D^3 + x^2D + 3x)P_5,$$

and

$$Dy = (x^2D^2 + 2xD)I_4 + (xD^3 - x^3D - 3x^2)I_5 + (xD^3 + x^3D + 3x^2)P_5.$$

Thus we have the result: the general solution of (38) is (44), where I_4 and I_5 are any invariants and P_5 is any pseudo-invariant with respect to σ . In this case the representation in Theorem 13 is seen to have a considerable advantage over that in Theorem 12.

UNIVERSITY OF MICHIGAN.

ON THE MINIMUM NUMBER OF POLYGONS IN AN IRREDUCIBLE MAP.*

By C. E. WINN.

In a recent paper Franklin¹ proved the number of polygons² in an irreducible map M to be at least 32. It is proposed here to shew with the help of certain new reductions that the number is at least 36.

Our main object is to set an upper limit on the number of pentagons touching a given polygon of M . When the contacts are consecutive, we use the fact that

A. *A polygon of 5, 6, 7 or $n > 7$ sides is reducible when in contact respectively with 3, 3, 4 or $n - 2$ adjacent pentagons.*³

When a pentagon has separate contacts with other pentagons, we note its reducibility⁴ if it touches the chain 5665. And, combining with A, we find

B. *A pentagon is reducible when in contact with 4 minor polygons of which the extremes are pentagons.*

Using the fact that

C. *A hexagon in contact with the chain 5565 or 55665 is reducible,*⁵ we shall prove a result analogous to B, namely

D. *A hexagon is reducible when in contact with 5 minor polygons of which the extremes are pentagons.*

This still allows the possibility of a hexagon of M touching two separate pairs of pentagons and two major polygons. But in this case we observe that

E. *A hexagon touching two separate pairs of pentagons is reducible when both pairs are in triad with another pentagon.*⁵

* Received June 24, 1938.

¹ "Note on the four color problem," *Journal of Mathematics and Physics*, vol. 16 (1938), p. 172 (published at Mass. Inst. of Technology).

² In an irreducible map every region is either a *minor* polygon of 5 or 6 sides, or a *major* polygon of more than 6 sides.

³ The only recent case is the third, given by the author, "On certain reductions in the four color problem," *Journal of Mathematics and Physics*, vol. 16 (1938), p. 159.

⁴ C. E. Winn, "A case of coloration in the four color problem," *American Journal of Mathematics*, vol. 49 (1937), p. 515.

⁵ *Loc. cit.*³. Unfortunately the claim in footnote 17 turns out to be unfounded.

As regards separate contacts with a heptagon, it is known that

F. *A heptagon touching 4 pentagons and 3 hexagons in any order is reducible.*⁵

We supplement this result by proving that

G. *A heptagon in contact with the chain 55655 is reducible.*

The details of the new reductions appear at the end of the paper, as well as those of a few configurations not employed here. The latter are as follows:

A pair of pentagons in triad with a heptagon and touching no other major polygon.

A pair of hexagons in contact with 55655 or 556655.

The next configuration is obtained by introducing into the ring of Errera⁶ the triad 575 occurring in a recent reduction of Franklin,¹ an *odd* number of hexagons being allowed.

Any ring formed of pairs 57 and, optionally, pairs 55 and hexagons in any order, the pairs 57 being oriented in one direction and each in triad with a pentagon that touches no other polygon of the ring.

As in Errera's case an isthmus in the reduced figure implies a Birkhoff ring in the original map when the ring encloses a single polygon, a pair or a triad—otherwise it may invalidate the result. Simple instances of unrestricted reducibility are those of 5(5)7666 about a pentagon and 5(5)765(5)76 about a hexagon, the digit in brackets denoting the pentagonal 'cap.' The final configuration is a modification of this type, namely

The ring 5(5)7665 about a pentagon.

If a_5 be the number of pentagons A_5 in M , and j_{5n} be the number of their contacts with higher polygons A_n , we shall have

$$(1) \quad j_{56} + 2 \sum_{n \geq 7} j_{5n} \geq 4a_5.$$

For the contribution of A_5 to the left member (which cannot exceed 10) is at least 4 when it touches one or no other pentagon. Also in view of the reductions A , B and the fact that an irreducible pentagon touches at least

⁶"Une contribution au problème des quatre couleurs," *Bulletin de la Société mathématique de France*, vol. 53 (1925), p. 42.

one major polygon,⁴ the possible contacts of A_5 with more than one pentagon are $55N5N$, $55nnN$, $55nNn$ and $5nN5n$, where $n \geq 6$ and $N \geq 7$, as hereafter. So in general the contribution is seen to be at least 4.

Let us now denote by $A_5^{(r)}$ an A_5 contributing $4 + r$ to the left of (1). Then $a_5^{(r)}$ being the number of such pentagons, we get

$$(2) \quad j_{56} + 2 \sum_{n \geq 7} j_{5n} = 4a_5 + \sum_{r=1}^6 ra_5^{(r)}.$$

Further, let $A_n^{(r)}$ be an A_n touching r pentagons. Their number being $a_n^{(r)}$, it follows from the last case of A that

$$(3) \quad j_{5n} = \sum_{r=1}^{n-2} ra_n^{(r)}.$$

The combination of (2) and (3) leads to

$$\sum_{r=1}^4 ra_6^{(r)} + 2 \sum_{n \geq 7} \sum_{r=1}^{n-2} ra_n^{(r)} = 4a_5 + \sum_{r=1}^6 ra_5^{(r)},$$

whence, seeing that $a_n \geq \sum_{r=1}^{n-2} a_n^{(r)}$, we get

$$2 \sum_{n \geq 6} (3n - 17)a_n + a_6^{(3)} + 2a_6^{(4)} \geq 4a_5 + \sum_{r=1}^6 ra_5^{(r)} \\ + \sum_{n \geq 7} \sum_{r=1}^{n-2} (3n - r - 17)a_n^{(r)}.$$

Consequently, if we shew

$$(4) \quad a_6^{(3)} + 2a_6^{(4)} + 2a_7^{(5)} \leq \sum_{r=1}^6 ra_5^{(r)} + 2 \sum_{n \geq 7} \sum_{r=1}^{n-2} (3n - r - 17)a_n^{(r)},$$

the negative term, given by $n = 7$, $r = 5$, being omitted from the double sum, it will follow that

$$\sum_{n \geq 6} (3n - 17)a_n \geq 2a_5.$$

Then we shall obtain by Euler's relation, as required,

$$(5) \quad a_5 + \sum_{n \geq 6} a_n \geq 3a_5 - 3 \sum_{n \geq 7} (n - 6)a_n = 36.$$

It may be remarked incidentally that, if no two pentagons of an irreducible map are adjacent, then at least 18 pentagons touch 3 or 4 hexagons.⁷

In fact, denoting the number of pentagons required by a_5' and a_5'' respectively, we have, since the contracts j_{5n} are separate,

⁷ Cp. Reynolds, "On the problem of coloring maps in four colors," *Annals of Mathematics*, vol. 28 (1926), p. I.

$$\begin{aligned}\sum_{n \geq 7} [\tfrac{1}{2}n] a_n &\geq \sum j_{5n} \geq 3a_5 - a_5' - 2a_5'' \\ &= 36 + 3 \sum_{n \geq 6} (n-6) a_n - a_5' - 2a_5'',\end{aligned}$$

whence

$$a_5' + 2a_5'' \geq 36.$$

To establish (4), we shall set against $A_6^{(3)}$, $A_6^{(4)}$ respectively one or two polygons $A_5^{(r)}$, $A_n^{(r)}$ adjacent to them as *compensating elements* which contribute to the right-hand side; and against $A_7^{(5)}$ one or two such elements $A_5^{(r)}$. It will then be necessary to verify that the number of sources of a given element, after reckoning *twice* the source $A_7^{(5)}$ yielding a *single* element, is at most equal to the corresponding coefficient on the right of (4).

From C and D we deduce that a hexagon of M that makes separate contacts with 3 pentagons must touch at least two major polygons. Thus $A_6^{(3)}$ is bounded by either $5n5N5N$, $55N5Nn$ or $5N5nN5$. In the former two cases we take as our element the last pentagon a adjacent to $A_6^{(3)}$, which is bounded by $6NmmN$, where $m \geq 5$, as hereafter.

In the last case let $bcde$ be the last four polygons about $A_6^{(3)}$, and let f be the outside polygon touching de . Then, if $c > 6$, we choose the element $b(6NmmN)$. If $c = 6$ and $f = 5$ or N , we choose e , which, on account of A , is bounded by $65N5N$ or $65mNN$ respectively. Finally, if $c = f = 6$, our element is $d(66 \cdots 65)$. If d is an $A_7^{(r)}$, we infer from F that $r \leq 3$, so that $A_7^{(4)}$ does not appear as an element.

The contacts of $A_6^{(4)}$ are $55N55N$ in view of A and C . Moreover, we conclude from E that one of these pairs of pentagons g, g' are not in triad with a third pentagon nor, by A , with another hexagon, when g, g' are in chain with a third pentagon. We here select two elements, namely $g, g'(65NmN)$ or $(656nN)$.

On account of A and G the ring round $A_7^{(5)}$ is $555N55N$. If the fourth (or last) polygon is also an $A_7^{(5)}$, we take the two pentagons h, h' touching both $A_7^{(5)}$'s. These are both $A_5^{(2)}$'s, their contracts being $75N57$ by A . But, if there is no adjacent $A_7^{(5)}$, we take the extreme pentagon i of the first three which, in virtue of B , touches an outside major polygon. We have then a single element bounded by $75NmN'$, where N' is not an $A_7^{(5)}$.

In each of the above rings about an element we have placed first its source. Consequently, a polygon with such contacts may occur as an element as often as one of its adjacent polygons fits into the first place of the ring (allowing for a reversal). We have thus to examine the possible occurrences of $A_5^{(r)}$, where $1 \leq r \leq 5$, and of $A_n^{(r)}$, where $1 \leq r \leq n-2$ or 3 , according as n is greater than or equal to 7.

The incidence of $A_5^{(1)}$ is at

$$a, b(6N55N); e, g, g'(65N5N); e(655NN); g, g'(6566N).$$

There is no repetition here, since the two adjacent hexagons touching g or g' cannot come first in any of the four rings.

The incidence of $A_5^{(2)}$ is at

$$a, b(6N56N); e, g, g'(656NN); h, h'(75N57); i(75N6N').$$

We observe that this element occurs twice at most in the first two rings, which contain only two hexagons; also these rings are distinct from the last two. Now, by supposition, the last polygon of the third ring is an $A_7^{(5)}$, whereas the last in the fourth ring is not. Hence an element $A_5^{(2)}$ can only appear twice in the third ring and once in the last, but not in both. This yields altogether a maximum of 2 occurrences, counting that at i twice.

The incidence of $A_5^{(3)}$ is at

$$a, b(6N5NN) \text{ or } (6N66N); e(65NNN); i(75N6N').$$

This element occurs only once in the last ring, as the third polygon, being next to a hexagon, is not an $A_7^{(5)}$. It can then fall but once elsewhere, namely in the first ring. Thus the maximum amounts to 3, seeing that none of the first three rings contain more than 3 hexagons.

The incidence of $A_5^{(4)}$ is at

$$a, b(6NN6N); i(75NNN').$$

The N between N and N' not being an $A_7^{(5)}$, we infer that this element can only fall twice in the last ring, and not then in the first. Hence, as the first ring contains but two hexagons, the maximum here is 4.

Lastly, and $A_5^{(5)}$ is only to be found once, at $a, b(6NNNN)$, while $A_5^{(6)}$ does not occur at all. So altogether the number of pentagons compensating $A_6^{(3)}$, $A_6^{(4)}$, and $A_7^{(5)}$ is not in excess of the first sum on the right of (4).

The number of occurrences of $A_n^{(r)}$ at $d(66 \cdots 65)$ cannot exceed $n-r$, i. e. the number of hexagons touching d , which is at most equal to the coefficient of $a_n^{(r)}$ in (4), unless $n=7$, $r=2$ or 3 . Moreover, in the last two cases the largest number of hexagons coming third in the sequence 6566 (or second in 6656) is found by inspection to be 4 or 2 respectively, i. e. not more than the coefficient of $a_7^{(r)}$. This concludes the demonstration of (4), and so of (5).

We now reduce^{*} the cases of D not contained in A or C , namely a hexagon touching 56565 and 56665.

^{*} The scheme of reduction is that explained in *loc. cit.* ⁴. To accommodate a transposition in one line a comma is sometimes used to mean 'or,' when it could not mean

N56565. See Fig. 1.

(1) $d, f, h = 2$; $f = 3$ or $h = 2$, unless $dfh = 332, 342$: $u = 1$ (2) $dfh = 243$

12 e to g or i	$f = 3, g = 3, 4$	$u = 1$	$e = 2$
32 h to a or e	$i = 4, d = 1, 4$	$u = 4$	$h = 2$
34 b to f	$c = 1$	$u = 1$	$b = 4$
13 g to i	$h = 4$	$u = 1$	
13 g to d	$i = 3$	$u = 4$	$g = 3$
32 e to a	$b = 1, d = 4$	$u = 4$	$e = 3$; $c, g = 2, 3$
		$u = 3$	

(3) $dfh = 244$

43 f to b	$c = 2$	$u = 4$	$f = 3, h = 3, 4$
		(1)	

(4) $dfh = 223$

13 g to e or c	$f = 4, d = 2, 4$	(2), (1)	$g = 3$
34 g to b	$a = 1, h = 2$	$u = 2$	
34 g to i	$ghi = 424$		
(23 a to f or h	$i = 1, g = 4, 1$	$u = 1$	
23 d to f or h	$a = 3$	$u = 2$	$abcd = 3243^*$
24 b to i	$a = 1$	$u = 4$	$b = 4$
12 e to c or h	$d = 4$ or $g = 3$	$u = 2$	$e = 2$
23 c to a or h	$b = 1, i = 4, 1$	$u = 2$	$c = 3$
		$u = 4$)	$g = 4$
		$u = 1$	

(5) $djh = 224$

31 g to i	$h = 2$	$u = 1$	
31 g to e or c	$f = 4, d = 2, 4$	(2), (1)	$g = 3$
34 g to b	$a = 1$	$u = 2$	$g = 4$
41 g to i	$h = 2$	$u = 1$	
41 g to e or c	$f = 3, d = 2, 3$	$u = 1$	$g = 1$
		(4)	

(6) $dfh = 323$

13 g to e	$f = 4$	(1)	$g = 3$
34 i to g	$h = 2$	$u = 3$	$i = 4$; $b, d = 3, 4$
		$u = 1$	

'and.' Thus $a, b = 2, 3$ means $a = 2$ or 3 and $b = 2$ or 3 . Also, if a chain affects adjacent polygons, it suffices to note the change in one of them. It should be added that, unless otherwise pointed out, an isthmus in the reduction implies a Birkhoff ring in the original figure, as can be at once verified.

*The absence of a 2 3 chain from d to b allows $a = 3$, as just given.

(7) $dfh = 324$

14 g to e or c	$f = 3, d = 2, 3$	$u = 1$	$g = 4$
12 h to f or a	g or $i = 3$	$u = 1$	
12 h to c	$g = 3, d = 4$	$u = 1$	$h = 2$
23 a to d or f	$c = 4, e = 1, 4$	$u = 4$	
23 h to d or f	$g = 1, e = 4, 1$	$u = 1$	$a = h = 3$
42 i to g or b	h or $a = 1$	$u = 1, 4$	$i = 2$
41 g to e or c	$f = 3, d = 3, 2$	$u = 1$	$g = 1$
24 b to f	$c = 3$	$u = 2$	
24 b to i	$f = 4$	$u = 1$	$b = 4$
43 b to d or h	$c = 2$ or $i = 1$	$u = 2, 1$	$b = 3$
32 b to f	$c = e = 4$	$u = 3$	$b = 2; d, h = 3, 2$
		$u = 2$	

(8) $dfh = 332$

23 h to f	$g = 4$	$u = 3$	$h = 2; d, b = 2, 3$
		(1) or equivalent	

(9) $dfh = 342$

41 e to c or i	$d = 2$ or $h = 3$	$u = 1$	$e = 4$
12 f to h or c	g or $e = 3$	$u = 1$	$f = 2$
32 d to a or h	$c = 4, i = 1, 4$	$u = 4$	
32 f to a or h	$g = 1, i = 4, 1$	$u = 3, 1$	$d = 2, f = 3$
42 e to g or a	f or $b = 1$	$u = 1, 3$	$e = 2$
41 g to i or d	h or $e = 3$	$u = 1$	$g = 1$
13 g to i or c	h or $e = 4$	$u = 1$	$g = 3$
21 h to f	$g = 4$	$u = 1$	
21 h to c	$e = 1$	$u = 1$	$h = 1$
14 a to f	$g = 2, i = 3$	$u = 2$	$a = 4; c, h = 1, 4$
		$u = 1$	

(10) $dfh = 423$

14 i to g or e	$h = 2, f = 2, 3$	(1)	$i = 4$
24 f to d or i	e or $g = 3$	$u = 3, 1$	$f = 4$
43 f to b	$c = e = 2$	$u = 4$	$f = 3; d, h = 3, 4$
		$u = 3$	

(11) $dfh = 424$

14 g to e	$f = 3$	(1)	$g = 4$
43 g to b	$a = 1, h = 2$	$u = 2$	
43 g to i	$ghi = 323, d = 4, 3$	$u = 3$	$g = 3; d = 3, 4$
		$u = 1$	

N56665. See Fig. 2.

(1) $j = 3$ unless $dfh = 243$ (342)

$u = 1$

(2) $j = 4$ unless $dfh = 222, 244$ (332) $u = 1$

 (3) $j = 4, dfh = 222$

12 i to a	$j = 3$	$u = 1$	$a = 1$
42 j to d	$a = 3$	$u = 1$	
42 j to f	$i = g = 3, d = 4$	$u = 3$	$j = 2; h = 2, 4$
		$u = 2$	

 (4) $j = 4, dfh = 244$

14 c to e	$d = 3$	$u = 1$	$c = 4$
43 j to b or h	$a = 1$ or $i = 2$	$u = 4$	
43 j to f	$j = f = 3$		
(23 d to b	$c = 1$	$u = 1$	$d = 3; f = 3, 2$
		$u = 3)$	$j = 3$
23 d to b	$c = 1$	$u = 1$	$d = 3$
42 c to a or h	$b = 1, j = 3, 1$	$u = 1$	
42 c to f	$c = f = 2$		
(21 c to a or e	b or $d = 4$	$u = 1$	
21 c to i	$d = 4, h = 3$	$u = 1$	$c = 1$
		$u = 1)$	$c = 2$
34 b to d	$c = 1$	$u = 1$	
34 b to f	$d = 4$	$u = 1$	$b = 4; h, j = 3, 4$
		$u = 1$ or (3)	

G. NN55655. See Fig. 3.

 (1) $d = 3$ or $h = 2$ or $f = 4$ $u = 1$

 (2) $dfh = 444$

43 d to b	$c = 2$	$u = 4$	$d = 3; f, h = 4, 3$
		$u = 1$	

 (3) $dfh = 244$

13 i to g, e, c	$h = 2, \text{etc.}$	(1)	$i = 3$
23 d to b	$c = 4$	$u = 4$	$d = 3$
31 i to b or g	$a = 4$ or $h = 2$	$u = 1$	$i = 1$
		(1)	

 (4) $dfh = 243$

14 e to c or i	$d = 3$ or $h = 2$	(1)	$e = 4$
34 b to h or e	i or $c = 2$	$u = 2, 1$	$b = 4$
14 c to e or i	d or $a = 3$	$u = 3, 2$	$c = 4$
12 b to d or f	$c = 3; e = 4, 3$	$u = 3$	$b = 2$
24 i to g or b	$h = 1$ or $a = 3$	$u = 4$	$i = 4$
		$u = 4$	

A pair of pentagons in triad with a heptagon and bounded elsewhere by minor polygons. See Fig. 4.

The polygons g, i and f, j must be hexagons on account of A and B respectively. The case where h is also a hexagon has lately been reduced by Franklin.¹ The remaining configuration, where h is a pentagon, can be colored immediately in the present reduced figure:

- (1) $a = b = 2 \quad u = 2, v = 1$
- (2) $a = 2, b = 3 \quad u = 2, v = 1$, unless $d = 3, c = 4$; then $u = 1, v = 2$
- (3) $a = b = 3 \quad u = 3, v = 1$, unless $d = 2, c = 4$; then $u = 1, v = 2$
- (4) $a = 3, b = 2 \quad u = 2$ or $4, v = 1$
or 4 .

A pair of hexagons touching 55655 or 556655. See Figs. 5 and 6.

We may suppose that a hexagon of the chain forms a triad with the given pair, as otherwise we get the reduction C .

We can color both figures immediately by marking u, v, b with 1, 2, 3. Then either 3 can be used for c , or else we get a choice for e (similarly for d and f).

There is an obvious extension when u, v have $2k, 2l$ sides and touch $k - 2, l - 2$ pairs of pentagons.

The reduction of the modified Errera ring R together with its pentagonal caps is made by suppressing these except for alignments¹⁰ connecting the remaining free vertices of caps, pairs of pentagons, hexagons and heptagons (see Fig. 7).

As to the coloration of R , we note that, just as in Errera's case a hexagon or pair of pentagons can always be colored when one neighboring polygon of R is already marked. Moreover, if the polygon a following cb (57) with the cap d is already marked, we can always color bdc . For b is adjacent to two other colors, namely that next to the part of R including c and the other color bounding the unreduced alignment L of b . The marking of b leaves 3 colors next to d . Then likewise 3 next to c .

We now have two possibilities according as the unreduced polygon e abutting d bears the latter color bounding L or not. If so, we mark c with this color and fill in R going away from b , with a final choice for bd . But, failing this combination at *any* pair 57, all such pairs can be colored in the *reverse* direction when the polygon of R previous to the pentagon is already marked. Consequently we can then fill in R , starting from b with the color of e , passing through a and finishing with a choice for cd .

¹⁰ The alignment crossing R at a pair of pentagons passes along their common side. Those crossing R at a hexagon or heptagon are kept apart by tracing them round the perimeter in the same sense from a given side of R .

We observe that in the reduced figure alignments crossing R may divide it into a number of parts. So an isthmus is formed at an alignment if it is the only one to cross it. The other possibilities of an isthmus are

- (1) if a polygon u makes two separate contacts with the same part of R , one contact only being reduced.
- (2) if two adjacent polygons v, w make separate reduced contacts with the same part of R .
- (3) if a polygon t makes separate reduced contacts with consecutive parts of R .

When R encloses a single polygon x , no alignment crosses it. In case (1) we get a 4-ring formed by u , two polygons of R and x , or a 5-ring about more than one polygon including a cap. In case (2) a similar 5-ring is formed by v, w , two polygons of R and x .

When R encloses a pair or a triad, it is crossed by 2 or 3 alignments respectively, or else two free vertices belonging to the same polygon or pair of pentagons on this side of R give rise to a polygon of 4 sides or less. Cases (1) and (2) yield the same result as above for the part of R considered. Lastly, in case (3) we have a 5-ring about more than one polygon formed by t , two polygons in the two parts of R and two of the enclosed polygons. Thus the reducibility is unrestricted for the configurations in question.

5(5)7665. See Fig. 8.

(1) unless $d = e = 2$ $u = 1$

(2) $d = e = 2$

24 d to a	$c = 3$	$u = 1$	$d = 4$
		(1)	

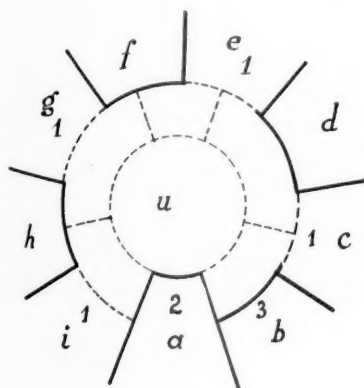


Fig. 1.

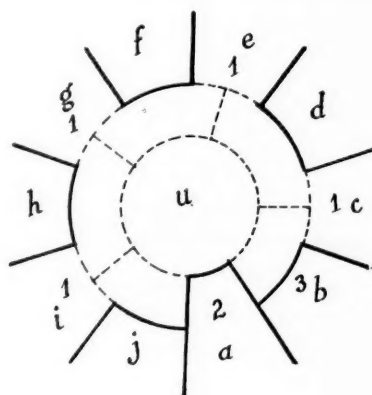


Fig. 2.

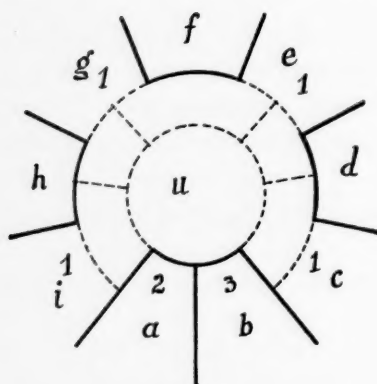


Fig. 3.

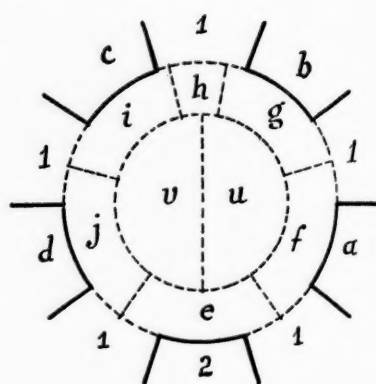


Fig. 4.

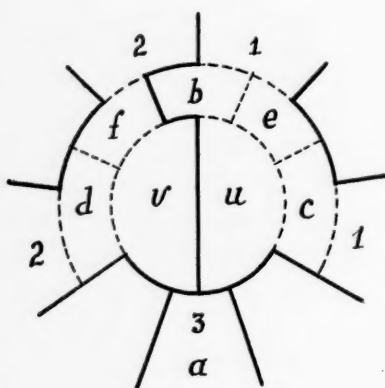


Fig. 5.

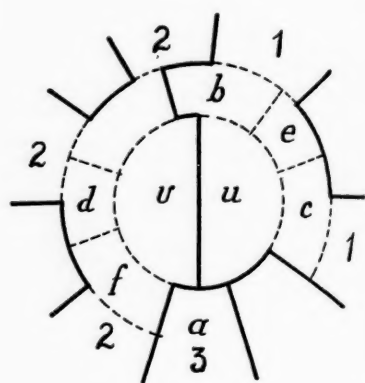


Fig. 6.

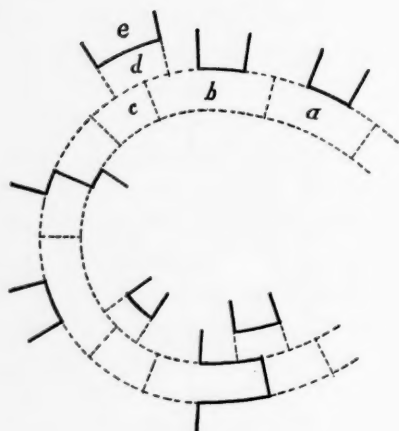


Fig. 7.

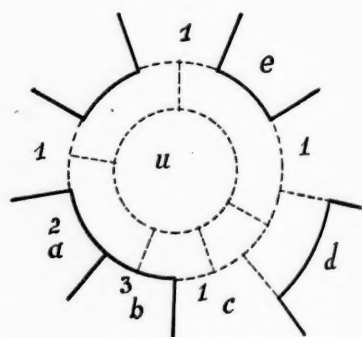


Fig. 8.

INFINITE PRODUCT MEASURES AND INFINITE CONVOLUTIONS.*

By E. R. VAN KAMPEN.

Introduction. The purpose of this paper is a systematic study of certain measurable functions on an infinite product space carrying a Lebesgue measure of the product type, especially the convergence theory of sequences of such functions and their distribution theory. Such a study is necessitated by the fact that these topics were considered during the last two decades from many different points of view by many authors and the development of the theory was quite slow. This warrants a uniform treatment of the central phases of the subject. An attempt will be made to approach each point by the method through which it is most easily accessible. There will result in this manner not only a systematical presentation of the general theory but quite naturally also several results which are not to be found in the literature.

Although some references are given, no attempt has been made at a serious historical study of the subject. Numbers in square brackets refer to the list of references at the end of the paper. References to statement numbers in parentheses are preceded by the roman numeral of the Part in which the statement occurs, except if the reference occurs in the same Part.

Part I concerns the theory of a product measure in a product space. The idea of such a measure developed from the theory of probability, cf. [1]. Later it took the form of a measure in certain special product spaces defined by means of a measure preserving mapping, cf. [27], pp. 496-497, [2 bis], [28], [29], [25], [3], [9]. Finally it took the form of a product of measures in a product space, defined directly as the product of given abstract measures in the factor spaces, cf. [20], [22], [8]. In Part I only so much is stated as is necessary for the understanding of what follows. A proof is given of the 0 — 1-theorem stated as I (6). The development of this theorem may be followed through a wide range of papers, for instance, [1], [28], [16], [21], [20], [9], [31].

A measurable function on one factor of the product space may be considered as a measurable function on the product space which is independent of all but one of the coördinates of each point of the product space. Part II concerns formal series of such functions, each series containing one term for each factor of the product space. The convergence theory of such series is

* Received June 19, 1939; Revised January 18, 1940.

easily accessible on the basis of the product measure introduced in Part I. The central theorem is the Three Series Theorem (Theorem I), which contains necessary and sufficient conditions for the convergence of the type of series in question. This theorem is due to Kolmogoroff, [17] and [18], who was led to this problem in connection with a particular series of independent functions introduced by Rademacher, [26].

In Part III it is shown how a simple mapping may transform a sequence of independent functions in the sense of Kolmogoroff, [20], or equivalently in the sense of Steinhaus, [11], into a sequence of functions of the type considered in Part II. Thus one can write, corresponding to every theorem of Part II, a corresponding theorem on series of independent functions. The mapping of the space of Part III on the product space of Part II is not a correspondence between points of these spaces, but a measure preserving correspondence between sufficiently extensive classes of measurable sets in these spaces. For considerations of the type used here such a correspondence is sufficient (cf. [32], § 3). The last paragraph of Part III contains the negative answer to a question of Kac and Steinhaus, (cf. [30], § 6).

Part IV concerns the convergence theory of infinite convolutions. The results of Part II are transferred to the theory of infinite convolutions by means of Theorem V, cf. [10], Theorem 32. A first proof of this theorem is based on Theorem IV in § 10 and IV (6) in § 17. Theorem IV, which is due to Jessen and Wintner ([10], pp. 84 and 85) is proved here by a method of Marcinkiewicz and Zygmund ([24], p. 119). The other result, IV (6), is usually proved by means of the theory of Fourier transforms ([10], Theorem 1). It is shown here that completely elementary methods are sufficient. A second proof of Theorem V is based on II (17) and is independent of IV (6). On the basis of Theorem V a list of theorems is stated without proof in § 20 and § 21. The relation of Parts II and IV is much more complicated than the relation of Part II and Part III. For instance, it can hardly be said that Theorems I and VI are equivalent, even though their analogy is immediately obvious; Theorem VI is due to Jessen and Wintner, [10], Theorem 34. Similarly, Theorem 3 of [15] corresponds to (and is used to prove) that part of the last statement of § 11 which has so far been proved. It would be desirable to invert this process. In other words, a simple proof of the last statement of § 11 would lead to a shorter proof and a better understanding of Theorem 3 of [15].

The pure theorem (Theorem VIII of § 22) is a generalization of (17) which is due to Jessen and Wintner ([10], Theorem 35). The remark that (17) may be extended to cover the case of any Hausdorff measure was communicated to me by Wintner. It may be of interest to investigate how far

one may allow more general given pure functions σ_n in Theorem VIII. It is, for instance, obvious that if $\star\sigma_n$ is convergent and σ_n is absolutely continuous for at least one value of n , then $\star\sigma_n$ is absolutely continuous.

It may be considered undesirable to prove a statement on convolutions like Theorem VI by means of series on infinite product spaces. However, at present it does not seem possible to prove Theorem VI without leaving the domain proper of distribution functions and their convolutions. Thus, for instance, the proof of Theorem VI which is sketched in § 20 includes a short excursion to the domain of Part I and an essential use of the theory of Fourier-Stieltjes transforms of distribution functions. An account of the latter may be found in [5] and many applications in [6], [10], [33], [35], [37]. However, in view of the criterion in IV (1) for the convergence of a sequence of distribution functions, it may be considered probable that eventually a reasonably simple proof of Theorem VI within the domain proper of that Theorem will be constructed. A presentation of the theory of distribution functions as a whole may be found in a course of lectures by Wintner at the Institute for Advanced Study, 1937-1938; a previous presentation is contained in [10].

The functions of Part II are real valued and the distributions of Part IV are 1-dimensional. This restriction is quite unessential. The extension to vector valued functions and more dimensional distributions involves only formal complications, but no essential difficulties. For such extensions in different situations compare [5], [6], [7], [10].

The convergence theory of series of independent random variables is not discussed in this paper. This theory, which from the historical point of view precedes the others, represents from the methodical point of view, an attempt to combine the advantages of the other theories. Apparently this combination succeeds only at the cost of some clarity, so that it seems preferable to consider the points of view of functions of independent variables and of distribution functions separately. A comprehensive treatment of this side of the question may be found in the well known treatise of P. Lévy: *Théorie de l'addition des variables aléatoires*, Paris (1937).

PART I. Product Measures.

1. Let $X_n, n = 1, 2, 3, \dots$ be an infinite sequence of sets and $X = \Pi X_n$ the product set of the X_n , i. e., the set of elements

$$(1) \quad x = \{x_n\} = \{x_1, x_2, x_3, \dots\},$$

where the n -th coördinate x_n of x is an arbitrary element of X_n . This product satisfies the commutative and associative laws with regard to any form of

permutation and bracketing of the sequence of integers $n = 1, 2, 3, \dots$. If this sequence is divided into two parts I, II, and correspondingly one writes $X = X_I \times X_{II}$, and if C_I is any subset of X_I , then $C_I \times X_{II}$ will be denoted by $(C_I)_X$. As an example for this convenient notation one has $(A_1 \times A_2)_X = A_1 \times A_2 \times X_3 \times X_4 \times \dots$, if $A_1 \subset X_1, A_2 \subset X_2$. The symbols X_n, X_n will be used to denote the products $X_1 \times X_2 \times \dots \times X_n, X_{n+1} \times X_{n+2} \times \dots$, so that $X = X_n \times X_n$; subsets of X_n, X_n will be provided with similar subscripts.

2. Let every space X_n carry an absolutely additive non-negative measure $\mu_n B_n$, defined for the sets B_n belonging to the field \mathfrak{B}_n of μ_n -measurable sets, and suppose that $\mu_n X_n = 1$. By the definition

$$(2) \quad \mu B = \prod_{k=1}^n \mu_k B_k, \text{ where } B = \left(\prod_{k=1}^n B_k \right)_X \text{ and } B_k \subset \mathfrak{B}_k,$$

and by subsequent uniquely determined extension (based, for instance, on the method of the exterior measure), an absolutely additive, non-negative measure μB may be defined for all sets B in a field \mathfrak{B} of μ -measurable subsets of X . This product measure $\mu = \prod \mu_n$ has the property that

(3) *The set $\prod A_n$, where $A_n \subset X_n$, is μ -measurable if and only if either each A_n is μ_n measurable, in which case $\mu \prod A_n = \prod \mu_n A_n = \prod \mu(A_n)_X$, or $\mu \prod A_n = 0$.*

Thus the use of μ both in (2) and as a notation for the product measure $\prod \mu_n$ is justified. In particular (3) implies that $\mu X = 1$. The proof of the above statements may be based, for instance, on the following intuitive lemma, used for this purpose by v. Neumann in a course of lectures at the Institute for Advanced Study, 1934-35.

(4) *If $A_n \subset \mathfrak{B}_n$, $A = \prod A_n$ and $A \subset \sum B^m$, where each B^m is a set of the type occurring in (3), then*

$$\prod \mu_n A_n \leq \sum \mu B^m$$

holds for the function μ defined in (2).

3. The measure $\mu = \prod \mu_n$ satisfies the commutative and associative laws with regard to any permutation and bracketing of the sequence of integers $n = 1, 2, 3, \dots$. In particular, if again $X = X_I \times X_{II}$, then $\mu = \mu_I \mu_{II}$, where μ_I, μ_{II} are the product measures of the μ_n belonging to X_I, X_{II} . Thus the theorem of Fubini may be applied to any factorization of X into two factors. This proves, for instance, the following statement, if one considers that the product $B_n \times C_n$ is equal to the common part of $(B_n)_X$ and $(C_n)_X$.

(5) *If $B_n \times C_n$ are measurable sets in X_n, X_n respectively, then*

$$\mu B_n \times C_n = \mu (B_n)_X (C_n)_X = \mu (B_n)_X \mu (C_n)_X.$$

The following well known theorem is important for the development of the theory of the product measure $\mu = \Pi \mu_n$:

(6) If a measurable set $C \subset X$ is of the form $(C_n)_X$ for every n , then either $\mu(C) = 1$ or $\mu(C) = 0$.

The condition concerning C is to the effect that a point (1) in C remains in C if any one of its coördinates is modified arbitrarily. The proof of (6) proceeds as follows:

A totally additive, non-negative measure function νB may be defined on B by the equation $\nu B = \mu BC$, where BC denotes the common part of B and C . If B is a set of the type $B = (B_n)_X$ then νB is, by (5), of the form

$$\nu B = \mu BC = \mu(B_n)_X (C_n)_X = \mu(B_n)_X \mu(C_n)_X = \mu B \mu C.$$

Since the measure ν is uniquely determined on B , by its values on all sets of the form $B = (B_n)_X$, this implies $\nu B = \mu BC = \mu B \mu C$ for every measurable set. On placing $B = C$ one obtains the statement (6).

4. The following convention will be used concerning X_n, μ_n, X, μ, f_n :

(7) X_n is a space which carries a measure μ_n such that $\mu_n X_n = 1$. The product space $X = \Pi X_n$ carries the product measure $\mu = \Pi \mu_n$, so that $\mu X = 1$. The symbol f_n represents a given μ_n -measurable function $f_n = f_n(x)$ on X , and the same symbol represents the μ -measurable function $f_n = f_n(x)$ on X , which is defined by: $f_n(X) = f_n(x_n)$ if the n -th coördinate of x is x_n , cf. (1).

Thus, for instance, in the symbol Σf_n , the f_n are thought of as functions on X , since otherwise addition would not have a meaning, and one finds for the k -th moment $M_k(f_n)$ of f_n the two expressions

$$M_k(f_n) = \int_{X_n} f_n(x_n)^k dX_n = \int_X f_n(x)^k dX,$$

if at least one of the integrals exists. The flexibility in the manipulation of integrals on X which one attains by means of Fubini's theorem is illustrated by the following example, which is typical of many situations in Part II:

(8) If $n \leq l < m$, $f_n(x_n)$ and $f_m(x_m)$ are integrable, and C is a set of the type $C = (C_i)_X$, then

$$\int_C f_n(x) f_m(x) dX = \int_C f_n(x) dX \int_X f_m(x) dX.$$

A special case of (7) is represented by (9). The use of the r_n on Z is equivalent with the use of the well-known Rademacher functions, cf. [26].

(9) Z_n is a space which consists of two points, Z'_n, Z''_n , each of which has the ν_n -measure $\frac{1}{2}$, so that $\nu_n Z_n = 1$. The product space $Z = \Pi Z_n$ carries the

measure $\nu = \Pi \nu_n$, so that $\nu Z = 1$. The function r_n is defined on Z_n by $r_n(Z'_n) = 1$, $r_n(Z''_n) = -1$; and r_n is defined on Z according to the last part of (7).

Two functions f on X and g on Y are said to be *equimeasurable* if $\mu[f(x) < \omega] = \lambda[g(y) < \omega]$ for every real ω . Here $[f(x) > \omega]$, for instance, represents the x -set defined by the inequality $f(x) > \omega$. Let the functions g_n on Y_n satisfy a convention similar to (7) and let f_n and g_n be equimeasurable for every n . If any limiting process (reducible to convergence in measure) is applied both to the f_n and the g_n , then the resulting functions are defined on sets of the same measure and equimeasurable on those sets.

A function f on X is said to be *symmetrically distributed* if $\mu[f(x) < \omega] = \mu[f(x) > -\omega]$ for every real ω . If the spaces X_n and Y_n are in 1-1 measure preserving correspondence, so that the same holds for X and Y , and if $f_n(x_n) = g_n(y_n)$ if the points x_n and y_n correspond, then it is easy to see that the function $f^*_n(x_n, y_n) = f_n(x_n) - g_n(y_n)$ is symmetrically distributed on $X_n \times Y_n$, hence also that $f^*_n(x, y)$ is symmetrically distributed on $X \times Y = \Pi(X_n \times Y_n)$. Moreover, $M_1(f^*_n) = 0$ and $M_2(f^*_n) = \bar{M}_2(f^*_n) = 2\bar{M}_2(f_n)$. Here $\bar{M}_2(f)$ denotes the value of the second moment of $f - M_1(f)$, i. e., the minimum value of the second moments of the functions $f - \text{const.}$; thus $\bar{M}_2(f) = M_2(f) - (M_1(f))^2$.

It is also easy to see that if f_n is symmetrically distributed on X_n , then the functions f_n on X_n and $r_n f_n$ on $X_n \times Z_n$ are equimeasurable, cf. (9). Thus, if f_n is symmetrically distributed for every n , then any limiting process applied to the sequences $\{f_n\}$ on X and $\{r_n f_n\}$ on $X \times Z = \Pi(X_n \times Z_n)$ leads to equimeasurable results.

PART II. Series of Functions of Independent Variables.

In Part II, a number of more or less known criteria are given for the convergence of a series $\sum f_n$ on $X = \Pi X_n$, where each $f_n = f_n(x)$ is obtained from a function $f_n = f_n(x_n)$ on x_n according to the convention at the beginning of § 4. The main criterion is stated as the "three series theorem" (Theorem I). It was obtained first by Kolmogoroff, who used the language of the theory of random variables, and restricted these to take on an at most enumerable set of values. His original proof in [17] turned out to be most convenient for a systematic presentation of the subject (§ 5 and § 6). Some simplification could be obtained. For instance, II (8) is replaced by II (2), and the proof of II (6) is separated into two parts, the first of which proves the separate lemma, II (5). The presentation of the proof of Theorem I may be varied in numerous ways.

In § 7 and § 8 conditions are given for the unconditional and absolute convergence of Σf_n on X . They follow easily from Theorem I. Theorem III is essentially simpler in character than Theorem I, from which it is here obtained, and may be proved directly by means of a lemma analogous to (5), concerning absolute convergence. Conditions as used in (12) of § 9 occur for $q = 2$, $p = 4$ in [11], Theorem 4, and for $q = 1$, $p = 2$ in [23], § 3. Theorem IV of § 10 is essential for the coördination of Part II and Part IV.

5. This § 5 contains some lemmas needed in the proof of the three series theorem. The convention I (7) is of course essential.

(1) *The series Σf_n is either almost everywhere convergent or almost everywhere divergent.* This is an immediate consequence of I (6).

(2) *If for an arbitrary $K > 0$ and every n , the function f'_n is defined by $f'_n = f_n$ or $f'_n = K$, according as $|f_n| \leq K$ or $|f_n| > K$, and also if f'_n is defined by $f'_n = f_n$ or $f'_n = -K$, according as $|f_n| \leq K$ or $|f_n| > K$, then Σf_n and $\Sigma f'_n$ are simultaneously almost everywhere convergent on X and almost everywhere divergent on X .* This is clear, since, for a fixed x , the passage from either of the series $\Sigma f_n, \Sigma f'_n$ to the other involves only terms which are of absolute value not less than K .

(3) *If $\Sigma M_2(f_n) < +\infty$ and $M_1(f_n) = 0$ for every n , then Σf_n is convergent almost everywhere on X .*

Suppose that $M_2(f_n)$ exists and $M_1(f_n) = 0$ for every n , and that Σf_n is divergent on a subset of positive measure of X . It will be shown that $\Sigma M_2(f_n)$ is divergent.

If s_n denotes the n -th partial sum of Σf_n , then the sequence $\{s_n\}$ is divergent on a subset of positive measure of X . Thus at almost every x on X , the oscillation of the sequence s_{m+1}, s_{m+2}, \dots is not less than a positive number $a = a(x)$ which depends on x but not on m . Thus it is evident that, for sufficiently small $b > 0$, the set D^* defined by the inequality $a = a(x) > b$ is not a zero set. This means that there exists a number $b > 0$ such that D is not a 0-set, where D is defined by the condition that $|s_n - s_m| > b$ holds for every m and at least one $n = n(x) > m$. If m is arbitrarily fixed, and $k > m$, let D^k be the set defined by $|s_n(x) - s_m(x)| \leq b$ for $m < n < k$ and $|s_k - s_m| > b$. The sets D^k are disjoint, ΣD^k contains D and D^k is a set of the type $(A, k)_X$. Let n be fixed in such a way that $\mu \sum_{m < k \leq n} D^k > \frac{1}{2} \mu D$. Since $M_1(f_k) = 0$, one sees from I (8) that

$$\int_{D^*} (s_n - s_k)(s_k - s_m) dX = \int_X (s_n - s_k) dX \cdot \int_{D^k} (s_k - s_m) dX = 0.$$

Taking in account the definition of D^k , one finds for $m < k \leq n$,

$$\begin{aligned} \int_{D^k} (s_n - s_m)^2 dX &= \int_{D^k} ((s_n - s_k)^2 + (s_k - s_m)^2) dX \\ &\geq \int_{D^k} (s_k - s_m)^2 \mu dX > b^2 \mu D^k, \end{aligned}$$

so that

$$\begin{aligned} \sum_{k=m+1}^n M_2(f_k) &= \int_X (s_n - s_m)^2 dX \\ &\geq \sum_{k=m+1}^n \int_{D^k} (s_n - s_m)^2 dX > \sum_{k=m+1}^n b^2 \mu D^k > \frac{1}{2} b^2 \mu D. \end{aligned}$$

Since m was arbitrarily chosen, it follows that $\sum M_2(f_n)$ is divergent, so that the proof of (3) is complete.

(4) If $\sum \bar{M}_2(f_n) < +\infty$, then $\sum f_n$ is convergent almost everywhere on X if and only if $\sum M_1(f_n)$ is convergent; in which case $\sum f_n$ converges in the mean (L^2) on X to $f = \sum f_n$ and

$$M_1(f) = \sum M_1(f_n), \quad \bar{M}_2(f) = \sum \bar{M}_2(f_n).$$

In fact, if g_n is defined by $g_n = f_n - M_1(f_n)$, then $M_1(g_n) = 0$ and $0 \leq M_2(g_n) = \bar{M}_2(f_n)$, so that $\sum M_2(g_n) < +\infty$. Thus according to (3), the series $\sum g_n$ is convergent almost everywhere on X . The first part of (4) is now evident from the definition of g_n . In order to prove the remaining statements of (4), put $g = \sum g_n = f - \sum M_1(f_n)$ and let t_n denote the n -th partial sum of $\sum g_n$. Since the functions g_n are orthogonal on X , one has $M_2(g) \geq M_2(t_n) = \sum_{k=1}^n M_2(g_k)$. From the equality one obtains by means of the theorem of Fatou the inequality $M_2(g) \leq \sum M_2(g_n)$, so that $M_2(g) = \sum M_2(g_n)$. Since the last identity may be read as the Parseval identity of the expansion $g = \sum g_n$, the remaining statements of (4) are now evident.

(5) If $|f_n(x)| < K$ for every x and every n , if $M_1(f_n) = 0$ for every n and if $\sum f_n$ is almost everywhere convergent, then $\sum M_2(f_n)$ is convergent.

Let $s_n = s_n(x)$ denote the n -th partial sum of the series $\sum f_n(x)$. Then the last assumption of (5) implies, in view of Egoroff's theorem, that $|s_n(x)| < N$ for some $N > 0$ and for every x in a set C of positive measure. Let the set C^n be defined by the inequalities $|s_k(x)| < N$ for $0 < k \leq n$, then C^n is a set of the form $C^n = (A_n)_X$. Moreover, $C^{n-1} \supset C^n \supset C$. On placing $T_n = \int_{C^n} s_n^2 dX$ and $D^n = C^{n-1} - C^n$, one has

$$T_k - T_{k-1} = \int_{C^{k-1}} f_k^2 dX + 2 \int_{C^{k-1}} f_k s_{k-1} dX - \int_{D^k} s_k^2 dX.$$

Application of I (8) gives the value $M_2(f_k)\mu C^{k-1}$ for the first integral and 0 for the second integral, since $M_1(f_n) = 0$ for every n . Thus, since $\mu C^{k-1} \geq \mu C$ and $|s_k| \leq K + N$ on D^k ,

$$T_k - T_{k-1} \geq M_2(f_k)\mu C - (K + N)^2\mu D^k.$$

Summation of this relation for $1 < k \leq n$ gives, since $\mu D^k = \mu C^{k-1} - \mu C^k$ and $\mu C^n \leq 1$,

$$K^2 \geq K^2\mu C^n \geq T_n \geq T_n - T_1 \geq \sum_{k=1}^n M_2(f_k)\mu C - (K + N)^2;$$

so that the convergence of $\Sigma M_2(f_n)$ is evident.

(6) If $|f_n(x)| < K$ for every x and every n , and if Σf_n is almost everywhere convergent on X , then both series $\Sigma M_1(f_n)$ and $\Sigma \bar{M}_2(f_n)$ are convergent; so that Σf_n belongs to the class (L^2) on X , (cf. (4)).

Let $f_n^*(x_n, y_n)$ be the function defined in § 4, so that $|f_n^*(x_n, y_n)| < 2K$, $M_1(f_n^*) = 0$, $M_2(f_n^*) = 2\bar{M}_2(f_n)$ and Σf_n^* is convergent almost everywhere on $X \times Y = \Pi(X_n \times Y_n)$. From (5), which is thus shown to be applicable, one obtains the convergence of $\Sigma \bar{M}_2(f_n)$. Since $\Sigma(f_n - M_1(f_n))$ is convergent almost everywhere by (4), it is clear that $\Sigma M_1(f_n)$ is convergent. Finally, that Σf_n belongs to the class (L^2) on X , is evident from (4).

6. From the statements (2) and (6) it is easy to obtain the three series theorem:

THEOREM I (Three series theorem). If A_n denotes the subset of X_n defined by the inequality $|f_n| \leq K$, then the series Σf_n is almost everywhere convergent on X if and only if all three series

$$(*) \sum \mu_n(X_n - A_n); (**) \Sigma \int_{A_n} f_n dX_n; (* *) \Sigma \left[\int_{A_n} f_n^2 dX_n - \left(\int_{A_n} f_n dX \right)^2 \right]$$

are convergent for a fixed $K > 0$, in which case they are convergent for every $K > 0$.

In fact, if f'_n is any one of the functions defined in (2) for $n = 1, 2, 3, \dots$, then $\Sigma f'_n$ is almost everywhere convergent on X , if and only if the same holds for Σf_n . Application of (6) to $\Sigma f'_n$ now shows that this is the case if and only if $\Sigma M_1(f'_n)$ and $\Sigma \bar{M}_2(f'_n)$ are convergent. Since clearly

$$\begin{aligned} M_1(f'_n) &= \int_{A_n} f_n dX_n \pm K\mu_n(X_n - A_n) \text{ and} \\ \bar{M}_2(f'_n) &= \int_{A_n} f_n^2 dX_n - \left(\int_{A_n} f_n dX_n \right)^2 \\ &\quad \pm 2K\mu_n(X_n - A_n) \int_{A_n} f_n dX_n + K^2(\mu_n(X_n - A_n) - \mu_n^2(X_n - A_n)^2), \end{aligned}$$

the convergence of $\Sigma M_1(f'_n)$ and $\Sigma \bar{M}_2(f'_n)$ for both functions f'_n is equivalent to the convergence of the three series occurring in Theorem I.

The following simple sufficient condition for convergence is an immediate consequence of Theorem I and may sometimes be easier of access; cf. [23], Theorem 5.

(7) *The series $\Sigma f_n(x)$ is convergent almost everywhere on X if so are both series $\Sigma M_1(f_n)$ and $\Sigma M_p(|f_n|)$ for a fixed p , $1 < p \leq 2$.*

In fact, if the second series is convergent, then so are the series $\Sigma \int_{X_n - A_n} f_n dX_n$, $\Sigma \mu_n(X_n - A_n)$ and $\Sigma \int_{A_n} f_n^2 dX_n$, where the A_n are the sets defined in Theorem I. Thus (7) is an immediate consequence of Theorem I.

On applying Theorem I to the functions f_n defined by $f_n(x) = 1$ or $f_n(x) = 0$ according as x does or does not belong to a given measurable set B_n on X_n , one obtains the following statement (which occurs as a lemma in the usual proof of Theorem 1; cf. [17], [18]).

(8) *If B_n is a measurable subset of X_n for every n , then the measure of the set of points x of X which have infinitely many coordinates X_n in the sets B_n is 0 or 1 according as $\Sigma \mu_n(B_n)$ is convergent or divergent.*

As remarked by M. Kac, application of Theorem I and of the other convergence criteria of Part II, to the series $\Sigma f_n/n$ leads to most of the well known sufficient conditions for the strong law of great numbers; cf. [11] p. 54. In fact, if $\Sigma f_n/n$ is convergent, then $(f_1 + \dots + f_n)/n \rightarrow 0$ as $n \rightarrow \infty$. For instance, on applying (3) to the series $\Sigma f_n/n$, one obtains the following sufficient criterion for the law of great numbers.

(9) *If $M_2(f_n)$ exists for every n , and if $\Sigma M_1(f_n)/n$ and $\Sigma M_2(f_n)/n^2$ are convergent, then*

$$(f_1 + \dots + f_n)/n \rightarrow 0, \text{ as } n \rightarrow \infty,$$

almost everywhere on X .

7. Since in (4), the series $\Sigma \bar{M}_2(f_n)$ has non-negative terms, and since the product space of X and its product measure μ satisfy the commutative law under any permutation of the factors X_n and μ_n , one may read (4) as follows:

(10) *If $M_2(f_n)$ exists for every n and if $\Sigma \bar{M}_2(f_n)$ is convergent, then the sum Σf_n is convergent almost everywhere, no matter in what fixed order the terms are taken, if and only if $\Sigma M_1(f_n)$ is absolutely convergent.*

If t' and t'' denote the sums of two rearrangements of Σf_n , then $t' = t''$ almost everywhere on X .

In order to prove this last statement, let g_n be defined again as $f_n - M_1(f_n)$, let t_n', t_n'' be the partial sums of the two rearrangements of Σg_n , and let T_n', T_n'' be the partial sums of the corresponding rearrangements of $\Sigma M_2(g_n)$. Then

$$\int_X (t_n' - t_n'')^2 dX = T_n' - T_n'',$$

so that the integral tends to 0, as $n \rightarrow \infty$. Hence, by Fatou's theorem,

$$\int_X (t' - t'')^2 dX = 0,$$

so that $t' = t''$ almost everywhere on X .

Using (10) instead of (4) in the proof, one obtains as in § 6 the following theorem, where notations of Theorem I are used:

THEOREM II. *The series $\Sigma f_n(x)$ is convergent almost everywhere on X , no matter in what fixed order the terms are taken, if and only if the series (*) and $\Sigma \int_{A_n} f_n^2 dX_n$ are convergent and the series (**) is absolutely convergent. Moreover, if t' and t'' denote the sum of two rearrangements of Σf_n , then $t' = t''$ almost everywhere on X .*

As an immediate consequence of Theorem II, one sees that if Σf_n is convergent almost everywhere on X , then constants a_n may be determined in such a way that the type of convergence of Theorem II holds for $\Sigma (f_n - a_n)$. One may select for instance as a_n the terms of the series (**) of Theorem I.

8. The type of convergence of Σf_n which was discussed in § 7 should be distinguished from the unconditional convergence of Σf_n almost everywhere on X . The latter is equivalent to the convergence of $\Sigma |f_n|$ almost everywhere in X , and obviously implies the type of convergence discussed in § 7. Simple examples show that the converse implication does not hold. For instance, the series $\Sigma r_n/n$ (cf. I (9)), satisfies the requirements of Theorem II, but $\Sigma |r_n|/n$ is divergent everywhere on Z .

A condition for the convergence of $\Sigma |f_n|$ almost everywhere on X may be obtained very easily from Theorem I. In fact, the set A_n of Theorem I is the same for f_n and $|f_n|$. Thus the statement of Theorem I for the series $\Sigma |f_n|$, is that this series is almost everywhere convergent on X if and only if the three series (*), $\Sigma [\int_{A_n} f_n^2 dx - (\int_{A_n} |f_n| dx)^2]$ and $\Sigma \int_{A_n} |f_n| dX_n$ are convergent. Since $\int_{A_n} f_n^2 dX_n \leq K \int_{A_n} |f_n| dX_n$, the result thus proved reduces to:

THEOREM III. *The series $\Sigma f_n(x)$ is absolutely convergent almost everywhere on X if and only if the two series*

$$\Sigma \mu_n(X_n - A_n), \quad \Sigma \int_{A_n} |f_n| dX_n$$

are convergent for a fixed value of $K > 0$, in which case they are convergent for every $K > 0$.

The simple sufficient conditions for absolute convergence in (11) are not equivalent for any two distinct values of p , cf. [34].

(11) *If for any fixed p , $0 < p \leq 1$, the series $\Sigma M_p(|f_n|)$ is convergent, then Σf_n is absolutely convergent almost everywhere on X .*

In fact, if $\Sigma M_p(|f_n|)$ is convergent, $0 < p \leq 1$, then so are the two series of Theorem III, so that Σf_n is absolutely convergent almost everywhere on X .

9. The condition $|f_n| < K$ in (6) insures that the convergence of Σf_n almost everywhere on X implies the convergence of $\Sigma \bar{M}_2(f_n)$. In this section a different condition will be discussed which leads to the convergence of a certain moment series, if a series Σf_n of symmetrically distributed functions f_n . In the special case $q = 1$, $p = 2$ of (12), the restriction to symmetrically distributed functions is not needed, as shown by Marcinkiewicz and Zygmund, [23], § 3. The resulting theorem is given here for completeness as (13). The moment condition of (12) has the form of an inverted Hölder inequality; thus $c \leq 1$ in (12) by Hölder's inequality.

(12) *Suppose that $0 < q < p$ and $c > 0$. Let the functions f_n on X_n be symmetrically distributed and let $cM_p(|f_n|)^{q/p} \leq M_q(|f_n|)$. Then the convergence almost everywhere on X of Σf_n implies that $\Sigma M_p(|f_n|)^{r/p}$ is convergent, where $r = \text{Max}(q, 2)$.*

First it will be shown that $m_n \rightarrow 0$, where $m_n = M_p(|f_n|)^{1/p}$. If B_n, C_n are the x_n -sets $B_n = [|f_n| > am_n]$, $C_n = [|f_n| > m_n/a]$, where $a > 0$, then

$$m_n^p \geq \int_{C_n} |f_n|^p dX_n \geq m_n^p a^{-p} \mu_n C_n$$

hence $\mu_n C_n \leq a^p$. And, from Hölder's inequality

$$\left(\int_{C_n} |f_n|^q dX_n \right)^p \leq (\mu_n C_n)^{p-q} \left(\int_{C_n} |f_n|^p dX_n \right)^q \leq (a^{p-q} m_n^q)^p.$$

Since obviously $\int_{X_n - B_n} |f_n|^q dX_n \leq a^q m_n^q$, it follows that

$$\mu(B_n - C_n) \geq a^q m_n^{-q} \int_{B_n - C_n} |f_n|^q dX_n \geq a^q m_n^{-q} (c m_n^q - a^q m_n^q - a^{p-q} m_n^q).$$

Since $\mu(B_m) \geq \mu(B_n - C_n)$, one sees that $\mu(B_n) > \text{const.} > 0$ if a is selected sufficiently small. Now if $\limsup m_n > 0$ and if in Theorem I, one takes $K < a \limsup m_n$, then the series (*) of Theorem I is not convergent, since infinitely many of its terms are $> \text{const.} > 0$. Since this is contradiction with the assumption that Σf_n is convergent almost everywhere on X , it follows that $m_n \rightarrow 0$.

Let A_n be the x_n set $A_n = [|f_n| \leq 1]$. Then

$$\mu_n(X_n - A_n) \leq \int_{X_n - A_n} |f_n|^p dX_n \leq m_n^p,$$

so that Hölder's inequality implies

$$\left(\int_{X_n - A_n} |f_n|^q dX_n \right)^p \leq \mu_n(X_n - A_n)^{p-q} \left(\int_{X_n - A_n} |f_n|^p dX_n \right)^q \leq m_n^{pq}.$$

Hence the assumption $M_q(|f_n|) \geq cm_n^q$ implies that

$$\int_{A_n} |f_n|^q dX_n \geq cm_n^q - m_n^p,$$

where the right side is positive for sufficiently large n , since $m_n \rightarrow 0$ and $q < p$. For such n one now obtains from Hölder's inequality of $r = 2 \geq q$ and from the definition of A_n if $r = q \geq 2$:

$$\int_{A_n} f_n^2 dX_n \geq \left(\int_{A_n} |f_n|^q dX_n \right)^{r/q} \geq (cm_n^q - m_n^p)^{r/q} = m_n^r (c - m_n^{p-q})^{r/q}.$$

This implies the statement of (12) that Σm_n^r is convergent. For $m_n \rightarrow 0$, $c > 0$, $p > q > 0$, and the series $(**)$ of Theorem I reduces to the form

$$\Sigma \int_{A_n} f_n^2 dX_n, \text{ since } f_n \text{ is symmetrically distributed.}$$

In case $q = 1$, $p = 2$ the condition in (12) that f_n is symmetrically distributed may be replaced by the much weaker condition that $M_1(f_n) = 0$. In fact, if $M_1(f_n) = 0$ and $cM_2(f_n)^{\frac{1}{2}} \leq M_1(|f_n|)$ for some $c > 0$, then the symmetrically distributed functions f_n^* of § 4 satisfy $dM_2(f_n^*)^{\frac{1}{2}} \leq M_1(|f_n^*|)$ for a $d > 0$ which depends only on c , as shown in [23], p. 71. Taking in account that the condition $cM_2(f_n)^{\frac{1}{2}} \leq M_1(|f_n|)$ is homogeneous in f_n , so that it is possible to normalize f_n by placing $M_2(f_n) = 1$, one obtains one half of the following theorem of Marcinkiewicz and Zygmund, the other half of which is clear from (4).

(13) If f_n satisfies the conditions $M_1(f_n) = 0$, $M_2(f_n) = 1$ and $M_1(f_n)$, $c > 0$ for every n , then the series $\Sigma c_n f_n$ with constant coefficients c_n is convergent almost everywhere on X or divergent almost everywhere on X , according as Σc_n^2 is convergent or divergent, cf. [23], § 3.

10. In this § 10 applications are given of a method of Zygmund ([38], § 2). They will be useful to establish the relations of Part II and Part IV.

(14) Let the n -th partial sum of the series $\sum c_n r_n$ on $Z = \Pi Z_n$ be denoted by t_n , where the c_n are constants and the r_n are defined in I (9). If a subsequence of the sequence t_n is convergent almost everywhere on Z with reference to the measure $\nu = \Pi \nu_n$, then $\sum c_n^2$ is convergent. Thus $\sum c_n r_n$ is almost everywhere convergent on Z .

In fact, if a subsequence of $\{t_n\}$ is convergent almost everywhere on Z , then there exists a constant M , and a subset C of Z of positive measure, such that $|t_m(z)| < M$ holds on C for arbitrarily large values of n . Thus,

$$M^2 \nu C \geq \int_C t_m^2 dZ = \sum_{n=1}^m c_n^2 \nu C + 2 \sum' c_n c_l \cdot \int_C r_n r_l dZ,$$

if \sum' denotes summation over the range $1 \leq n < l \leq m$. Now the functions $r_n r_l$, $n < l$ are orthogonal to each other and orthogonal to a constant function on Z . Thus, the Parseval inequality, when applied to the characteristic function of C , shows that

$$\sum' \left(\int_C r_n r_l dZ \right)^2 \leq \nu C - (\nu C)^2 \leq \frac{1}{4},$$

so that, by Schwarz' inequality,

$$(2 \sum' c_n c_l \cdot \int_C r_n r_l dZ)^2 \leq 4 \sum' c_n^2 c_l^2 \cdot \sum' \left(\int_C r_n r_l dZ \right)^2 \leq \frac{1}{2} (\sum c_n^2)^2,$$

and finally

$$M^2 \nu C \geq \sum_{n=1}^m c_n^2 \cdot \left(\nu C - \frac{1}{\sqrt{2}} \right).$$

Since M and C may be selected in such a way that νC is arbitrarily near to 1, this completes the proof that $\sum c_n^2$ is convergent. And now (4) implies that $\sum c_n r_n$ is almost everywhere convergent on Z .

(15) Let s_n be the n -th partial sum of $\sum f_n$ on $X = \Pi X_n$. If the subsequence $\{s_{n_k}\}$ of $\{s_n\}$ is convergent almost everywhere on X , then there exist constants a_n such that $\sum (f_n - a_n)$ converges almost everywhere on X .

Consider first the case where the distribution of f_n on X_n is symmetric, i. e., where the sets defined by $f_n > \omega$ and $f_n < -\omega$ have equal measures for every real ω . Then the function $r_n f_n$ on $X_n \times Z_n$ and f_n on X_n are equimeasurable, i. e., the sets defined by $f_n > \omega$ on X_n and $f_n r_n > \omega$ on $X_n \times Z_n$ have equal measure for every real ω . Thus if t_n denotes the n -th partial sum of the series $\sum f_n r_n$ on $X \times Z = \Pi (X_n \times Z_n)$, then the sequence $\{t_{n_k}\}$ is con-

vergent almost everywhere on $X \times Z$. Thus, by Fubini's theorem, $\{t_n\}$ is convergent almost everywhere on Z at almost every fixed point of X . Hence, by (14), $\Sigma f_n r_n$ is convergent almost everywhere on Z at almost every fixed point of X . In view of Fubini's theorem, this implies that $\Sigma f_n r_n$ is convergent almost everywhere on X at almost every fixed point z of Z . Since the functions f_n on X_n and $f_n g_n$ on $X_n \times Z_n$ are equimeasurable for every fixed z_n in Z_n , this implies finally that Σf_n is almost everywhere convergent on X , so that one may choose $a_n = 0$ in the case under consideration.

Now let f_n be arbitrary and let f_n^* be the function introduced in § 4. Clearly, $\Sigma f_n^*(x, y)$ has the same convergence property on $X \times Y$ as Σf_n has on X . Moreover $f_n^*(x_n, y_n)$ has a symmetric distribution on $X_n \times Y_n$. Thus, by the case of (15) which has already been proved, $\Sigma f_n^*(x, y)$ is almost everywhere convergent on $X \times Y$. This implies that $\Sigma f_n^*(x, y)$ is almost everywhere convergent on X at at least one fixed point y_0 of Y . On placing $f_n(y_0) = a_n$, one obtains the statement of (15). This proof was obtained from a proof in [24] by a slight simplification.

THEOREM IV. *If Σf_n is convergent in measure on $X = \Pi X_n$, then Σf_n is almost everywhere convergent on X .*

In fact, if Σf_n is convergent in measure on X , then a classical theorem states that a subsequence of the sequence of partial sums of Σf_n is convergent almost everywhere on X . Thus, by (15), there exist constants a_m for which $\Sigma(f_n - a_n)$ is convergent almost everywhere on X . Since Σf_n is convergent in measure on X , one may choose the constants a_n equal to 0. Thus the proof of Theorem IV is complete.

11. The method of Zygmund which led in § 10 to (14) and (15) may be used to prove certain additional theorems, examples of which are given here.

Let γ_{mn} be real numbers such that $\gamma_{mn} \rightarrow 1$ as $m \rightarrow \infty$ for every n and let S_m denote the sum $S_m = \sum_{n=1}^{\infty} \gamma_{mn} f_n$, if it is convergent.

(16) *If a sequence of constants b_m exists such that the sequence $\{S_m - b_m\}$ is bounded on a set C of positive measure, then there exist constants a_n such that $\Sigma(f_n - a_n)$ is convergent almost everywhere on X , cf. [38], p. 97 and p. 100.*

First, one may suppose that $\gamma_{mn} = 0$ for every fixed m and $n > N = N_m$. This is obvious, since the series defining S_m are, by Egoroff's theorem, uniformly convergent on a subset of C which has positive measure.

Next, using the argument in the second part of the proof of (15), one sees that it is sufficient to consider the case in which the distribution of each

f_n on X_n is symmetric, choosing $b_m = 0$ for every m and $a_n = 0$ for every n . And the first part of the proof of (15) may be used to reduce the proof of (16) to the case where each f_n is of the form $c_n r_n$, cf. (14) and I (9).

Furthermore, on omitting a finite number of terms of the series $\Sigma c_n r_n$, one may suppose that the measure of the set on which the sequence $\{S_n\}$ corresponding to $\Sigma c_n r_n$ is bounded, is larger than $2^{-\frac{1}{2}}$. Finally, application of the method used in the proof of (14) establishes an upper bound for the sequence of numbers

$$\sum_{n=1}^{N_m} \gamma_{mn}^2 c_n^2.$$

Since $\gamma_{mn} \rightarrow 1$ as $n \rightarrow \infty$ for every m , this implies the convergence of Σc_n^2 , hence the convergence almost everywhere of $\Sigma c_n r_n$. This completes the proof of (16).

The generality of the summation method in (16) allows a large number of applications. However the next statement may not be obtained from (16), since the set C^m is allowed to depend on m . The notation S_m introduced at the beginning of this § 11 is used again.

(17) *For every $\epsilon > 0$, let there exist a constant $M = M_\epsilon$ and sets C^m on X , such that $\mu C^m > 1 - \epsilon$ and that $|S_m| < M$ holds on C^m for every m . Then there exists a sequence of constants a_n , such that $\Sigma(f_n - a_n)$ is convergent almost everywhere on X .*

The proof of (17) is not essentially distinct from the proof of (16). The measure of the set on which the method of (14) is applied may be selected larger than $2^{-\frac{1}{2}}$ by choosing ϵ sufficiently small. The case where S_m is selected to be the partial sum s_m of Σf_n will be used in § 19.

A comparison of (16) and (17) naturally suggests the truth of the following statement, of which no proof is known:

If $0 < a < 1$, $M > 0$ and if there exists a sequence of constants b_m and a sequence of sets C^m in X such that $|S_m - b_m| < M$ on C^m and $\mu C^m > a$ for every m , then there exists a sequence of constants a_n such that $\Sigma(f_n - a_n)$ is convergent almost everywhere on X .

In the particular case where each S_m is a partial sum of the series Σf_n , a proof may be obtained by comparing Theorem V of Part II and Theorem 3 of [15]. An analysis of the proof of this last theorem might lead to a proof of the above conjecture.

PART III. Series of Independent Functions.

Convergence criteria of series of independent functions on $[0, 1]$ in the sense of Kolmogoroff or, equivalently, in the sense of Steinhaus (cf. [20] and [11]), may be derived immediately from the convergence criteria of Part II for $\sum f_n$ on $X = \prod X_n$, where each f_n is a function on X_n . It is immaterial whether the independent functions are defined on the interval $[0, 1]$ or on any set Z carrying a measure ν such that $\nu Z = 1$.

12. Let $\{g_n(z)\}$, $\{g'_n(z)\}$ be two sequences of real valued measurable functions defined on spaces Z, Z' carrying abstract Lebesgue measures ν, ν' respectively. These sequences are said to be *equimeasurable* if for any finite number of Borel sets $\Omega_1, \dots, \Omega_k$ of real numbers, one has $\nu C = \nu' C'$, where C, C' are the sets defined by

$$C: g_n(z) \subset \Omega_n, (n=1, \dots, k); \quad C': g'_n(z') \subset \Omega_n, (n=1, \dots, k).$$

The functions $g_n(z)$ on Z are said to be *independent* on Z if for any finite number of Borel sets $\Omega_1, \dots, \Omega_k$ one has $\nu C = \prod_{n=1}^k \nu C_n$. Here C_n and C are defined by the inequalities:

$$C_n: g_n(z) \subset \Omega_n; \quad C: g_n(z) \subset \Omega_n \text{ for } n=1, \dots, k.$$

The proofs of the following statements are evident, cf. e.g., [24].

If $\{g_n\}$ and $\{g'_n\}$ are equimeasurable sequences, then the functions obtained from $\{g_n\}$ and $\{g'_n\}$ by any limiting process (reducible to convergence in measure) are equimeasurable functions.

If $\{f_n\}$ is a sequence of functions obtained on $X = \prod X_n$ by the convention I (7), then $\{f_n\}$ is a sequence of independent functions on X .

If the sequences $\{f_n\}$, $\{g_n\}$ are independent on X, Z respectively, then these sequences are equimeasurable, if and only if the functions f_n and g_n are equimeasurable for every n .

13. Now let $\{g_n(z)\}$ be a sequence of independent functions on a space Z , which carries a measure ν for which $\nu Z = 1$. For every n , let X_n be the space of real numbers. Let μ_n be defined on X_n by the definition $\mu_n(\Omega_n) = \nu(C_n)$, where Ω_n is any Borel set in X_n and C_n is the z -set $C_n = [g_n(z) \subset \Omega_n]$. Let a function $f_n(x_n)$ be defined on X_n by placing $f_n(x_n) = x_n$ for every (real number) x_n in X_n . From these definitions it is clear that g_n and f_n are equimeasurable functions. Now let the measure $\mu = \prod \mu_n$ be introduced on the product space $X = \prod X_n$. Then the statements of § 12 imply that the sequences $\{f_n\}$ on X and $\{g_n\}$ on Z are equimeasurable sequences of functions. Thus one obtains the following theorem:

(1) If $\{g_n\}$ is a given sequence of ν -measurable independent functions on a space Z for which $\nu Z = 1$, then there exists a sequence $\{f_n\}$ of μ -measurable functions, defined on a product space $X = \Pi X_n$ by means of the convention I (7), such that $\{g_n\}$ and $\{f_n\}$ are equimeasurable sequences.

Thus, according to § 12, the criteria of Part II for different types of convergence of series of the type of Σf_n , retain their validity if Σf_n is replaced by Σg_n , where the g_n form a sequence of independent functions on a space Z of total ν -measure 1.

As another application, note that as a consequence of (1), Theorem I, § 2 of [23] is a very special case of (16.1) on p. 280 in [9]. The more special character of the former, has as one of its consequences, that the former does while the latter does not allow a direct generalization to the case $p = 1$, cf. [24].

14. This § 14 contains the negative answer to a question of Kac and Steinhaus concerning the relation between completeness and independence, cf. [30], § 6.

If the functions g_n on Z are bounded, then clearly the sequence of powers $f_n^0, f_n^1, f_n^2, \dots$ is complete on X_n , so that the set of all monomials in the f_n is complete on $X = \Pi X_n$. It cannot be concluded that the set of all monomials in the g_n is complete on Z , or even that if this set is not complete, then the set of independent functions g_n on Z can be enlarged by a non-constant function. In fact, if $g(x)$ is defined on $[0, 1]$ by $g(x) = 2^{\frac{1}{2}}x^{\frac{1}{2}}$ or $1 - 2^{\frac{1}{2}}(1-x)^{\frac{1}{2}}$, according as $0 \leq x \leq \frac{1}{2}$ or $\frac{1}{2} < x \leq 1$, then the sequence $g^0, g^1, g^2, g^3, \dots$ is not complete on $[0, 1]$, and if f is a function on $[0, 1]$ such that f and g are independent, then f is a constant almost everywhere. If f is not constant and f and g are independent, let A be a set defined by $A: f(x) < \omega$, where ω is selected in such a way that the measure of A is neither 0 nor 1. A contradiction is now easily obtained by applying the fundamental theorem of the calculus to the characteristic function of A .

Closely related is the remark that a sequence of independent functions g_n (or even of their powers $(g_n)^r$) on $[0, 1]$ is never complete. In fact, if f_n is the function on X_n corresponding to g_n , and $k \neq l$, then $f_k f_l$ cannot be approximated by linear combinations of the $(f_n)^r$ unless either f_k or f_l is constant.

PART IV.

In this part certain convergence criteria for infinite convolutions will be derived from the corresponding criteria in Part II. As soon as the connection between the theory of infinite convolutions and the theory of the series in Part II has been established, one may transcribe these theorems without

further proof (although theorems of Part II and Part IV are never equivalent). This explains the list of theorems without proofs in § 20 and § 21.

15. A distribution function $\sigma = \sigma(t)$, $-\infty < t < +\infty$, is defined to be a monotone function such that $\sigma(-\infty) = 0$, $\sigma(+\infty) = 1$. Two distribution functions will be considered as *identical* if they are equal at their common continuity points. Thus two distribution functions are identical if they are equal at a dense set of values of t . A sequence $\{\sigma_n\}$ of distribution functions is said to converge if there exists a distribution function σ , such that $\sigma_n(t) \rightarrow \sigma(t)$ holds at every continuity point of σ . Note that $\{\sigma_n\}$ is not considered to be convergent if $\sigma_n(t) \rightarrow \alpha(t)$ holds for every t but $\alpha(t)$ is not a distribution function. If σ_n, σ are distribution functions, then $\sigma_n \rightarrow \sigma$ obviously holds if $\sigma_n(t) \rightarrow \sigma(t)$ for a dense set of values of t . The following criterion for the convergence of a sequence $\{\sigma_n\}$ of distribution functions is not quite obvious.

(1) *The sequence $\{\sigma_n\}$ is not convergent if and only if there exists an $a > 0$ and for every n , an $m > n$ such that $|\sigma_m(t') - \sigma_n(t'')| > a$ holds for every t' and t'' in at least one interval of length a .*

That the existence of such an a is not compatible with the convergence of $\{\sigma_n\}$ is obvious. Thus, it remains to prove that if such an a does not exist, then $\{\sigma_n\}$ is convergent.

By a theorem of Helly, a subsequence of $\{\sigma_n\}$ may be selected which tends everywhere to a monotone function $\alpha(t)$. If t_0 is a continuity point of α , and $\epsilon > 0$ is arbitrary, let $\delta > 0$ be chosen such that $|\alpha(t_0 \pm 2\delta) - \alpha(t_0)| < \epsilon$. By assumption, there exists an $n = n_{\epsilon\delta}$ such that for every $m > n = n_{\epsilon\delta}$, the inequality $|\sigma_m(t') - \sigma_n(t'')| \leq \epsilon$ holds for at least one set of values t', t'' on every interval of length δ .

Let the element σ_m , $m > n = n_{\epsilon\delta}$, of the subsequence which determined α be such that $|\alpha(t_0 \pm 2\delta) - \sigma_m(t_0 \pm 2\delta)| < \epsilon$, so that $|\sigma_m(t_0 \pm 2\delta) - \alpha(t_0)| < 2\epsilon$. Hence, the definition of $n = n_{\epsilon\delta}$ implies first $|\sigma_n(t_0 \pm \delta) - \alpha(t_0)| < 3\epsilon$, and then $|\sigma_p(t_0) - \alpha(t_0)| < 4\epsilon$ for every $p > n = n_{\epsilon\delta}$. Thus, $\sigma_n(t_0) \rightarrow \alpha(t_0)$ at every continuity point of $\alpha(t)$. Finally, α must be a distribution function. In fact, since $\alpha(+\infty)$ and $\alpha(-\infty)$ exist, the condition which was used above to determine δ as a function of ϵ and t_0 , may be satisfied by the same $\delta > 0$ for a fixed $\epsilon > 0$ and every t_0 which is sufficiently large. Since $\sigma_p(+\infty) = 1$ and $\sigma_p(-\infty) = 0$ for every p , this implies that $\alpha(+\infty) = 1$ and $\alpha(-\infty) = 0$; so that α is a distribution function, and the proof of (1) is complete.

A distribution function $\sigma = \sigma(t)$ determines uniquely a Lebesgue-Stieltjes measure on $-\infty < t < +\infty$. This measure, which will also be

denoted by σ , may be obtained by a well-known extension, from the definition $\sigma\Delta = \sigma(b) - \sigma(a)$, where Δ is the t -set $a \leq t < b$ and a, b are continuity points of σ . Thus σt denotes the σ -measure of the point t , i. e., the jump of $\sigma(t)$ at t . The integral of an integrable function $g(t)$ with respect to this measure σ is the Lebesgue-Stieltjes integral

$$\int_{-\infty}^{+\infty} g(t) d\sigma(t).$$

The *spectrum* of σ is defined to be the set of those values of t for which $|\sigma(t+\epsilon) - \sigma(t-\epsilon)| > 0$ holds for every $\epsilon > 0$. The *point spectrum* of σ is the set of those t for which $\sigma t > 0$, i. e., the set of discontinuity points of σ .

16. If σ_1 and σ_2 are two distribution functions, then the *convolution* $\sigma_1 * \sigma_2$ of σ_1 and σ_2 is defined by

$$\sigma_1 * \sigma_2(t) = \int_{-\infty}^{+\infty} \sigma_1(t-s) d\sigma_2(s) = \int \int_{r+s < t} d\sigma_1(r) d\sigma_2(s)$$

where the double Lebesgue-Stieltjes integral on the right represents the measure of the (r, s) -set $[r+s < t]$, if in the (r, s) -plane the product of the two measures σ_1 and σ_2 is used. From this second expression for $\sigma_1 * \sigma_2$ the following statement is clear.

(2) The function $\sigma_1 * \sigma_2$ is a distribution function and the point spectrum (spectrum) of $\sigma_1 * \sigma_2$ may be obtained by adding arbitrary elements of the point spectra (spectra) of σ_1 and σ_2 (and forming the closure). Moreover,

$$\sigma_1 * \sigma_2(t \pm 0) = \int_{-\infty}^{+\infty} \sigma_1(t-s \pm 0) d\sigma_2(s) \text{ and } \sigma_1 * \sigma_2 t = \sum_{r+s=t} \sigma_1 r \cdot \sigma_2 s.$$

Using Fubini's theorem, one sees that the convolution of any finite number of distribution functions satisfies the commutative and associative laws. In what follows $\sigma, \tau, \sigma_n, \tau_n$ always denote distribution functions.

(3) If $\sigma_n \rightarrow \sigma$ and $\tau_n \rightarrow \tau$ as $n \rightarrow \infty$, then $\sigma_n * \tau_n \rightarrow \sigma * \tau$.

It is sufficient to show that $\sigma_n * \tau_n(t_0) - \sigma * \tau(t_0) \rightarrow 0$, as $n \rightarrow \infty$, at every continuity point t_0 of $\sigma * \tau$. In fact, one obtains the statement $\sigma * \tau_n(t_0) - \sigma * \tau(t_0) \rightarrow 0$ on replacing σ, σ_n, τ_n by τ, τ_n, σ , and the two together imply (3).

If $\epsilon > 0$ is given, let r_1, \dots, r_p be a finite number of discontinuity points of $\sigma(t)$ which are such that the sum of the jumps at all remaining discontinuity points of $\sigma(t)$ is less than ϵ . Since, by (2), the p numbers $t_0 - r_k$

($k = 1, \dots, p$) are continuity points of $\tau(t)$, one may determine the non-overlapping intervals $I_k : a_k < t < b_k$, in such a way that $\Sigma_k[\tau(b_k) - \tau(a_k)] < \epsilon$, where $a_k < t_0 - r_k < b_k$, and the a_k, b_k are continuity points of $\tau(t)$. Next, one can determine M in such a way that $\Sigma_k[\tau_n(b_k) - \tau_n(a_k)] < 2\epsilon$ for $n > M$. On the other hand, one may determine $N > M$ in such a way that $|\sigma_n(t) - \sigma(t)| < 2\epsilon$ holds for every $n > N$ and for every t which is not in any of the intervals $t_0 - b_k < t < t_0 - a_k$. Thus, on separating the contributions of the intervals I_k and of the rest of the integration domain, one has

$$|\sigma_n * \tau_n(t_0) - \sigma * \tau_n(t_0)| = \left| \int_{-\infty}^{+\infty} [\sigma_n(t_0 - s) - \sigma(t_0 - s)] d\tau_n(s) \right| \leq 2\epsilon + 2\epsilon,$$

if $n > N$ ($> M$); so that the proof of (3) is complete.

Let ω denote the distribution function for which $\omega(t) = 0$ or 1 according as $t < 0$ or $t > 0$.

(4) If σ_n, τ_n are such that $\sigma_n \rightarrow \sigma$, $\sigma_n * \tau_n \rightarrow \sigma$ as $n \rightarrow \infty$, then $\tau_n \rightarrow \omega$.

If $\epsilon > 0$ is given and $t, -t$ are continuity points of σ for which $\sigma(t) - \sigma(-t) > 1 - \epsilon$, then N may be chosen such that $\sigma_n(t) - \sigma_n(-t) > 1 - 2\epsilon$ and $\sigma_n * \tau_n(t) - \sigma_n * \tau_n(-t) > 1 - 2\epsilon$ for $n > N$. Thus, for $n > N$,

$$\begin{aligned} 1 - 2\epsilon &< \int_{-\infty}^{-2t} + \int_{2t}^{+\infty} + \int_{-2t}^{2t} [\sigma_n(t-s) - \sigma_n(-t-s)] d\tau_n(s) \\ &\leq 2\epsilon + \tau_n(2t) - \tau_n(-2t), \end{aligned}$$

where the variation of τ_n in the first two integrals and the integrand of the last integral have been majorized by 1. Since $\tau_n(2t) - \tau_n(-2t) > 1 - 4\epsilon$ for any ϵ and a suitable $t = t_\epsilon$, one can find a subsequence of the sequence $\{\tau_n\}$ which tends to a distribution function τ . From (3) and the first assumption of (4) one sees that the corresponding subsequence of $\sigma_n * \tau_n$ tends to $\sigma * \tau$; so that $\sigma * \tau = \sigma$. Now if $\tau_n \rightarrow \omega$ did not hold, one could select τ to on any interval consisting of negative t -values. Thus $\tau = \omega$, as stated in (5).

(5) If $\sigma * \tau = \sigma$ then $\tau = \omega$.

Let $t_0 > 0$ be given; the maximum m of $\sigma(t + t_0 + 0) - \sigma(t - 0)$ for $-\infty < t < +\infty$ exists and is positive, and is attained at every point of a bounded closed t -set, A_0 . Let t_1, t_2 be the least and largest values of t in A_0 . Then (2) implies that

$$\sigma(t_1 + t_0 + 0) - \sigma(t_1 - 0) = \int_{-\infty}^{+\infty} [\sigma(t_1 + t_0 - s + 0) - \sigma(t_1 - s - 0)] d\tau(s) = m.$$

Since the total variation of τ is 1, and since $0 \leq \sigma(t + t_0 + 0) - \sigma(t - 0) \leq m$, this is possible only if $\tau(s)$ has the variation 0 on any interval on which $\sigma(t_1 + t_0 - s + 0) - \sigma(t_1 - s - 0) < m$ holds everywhere. In particular, the total variation of τ must be 0 on any interval consisting only of positive t -values. Using t_2 instead of t_1 , one sees that τ also has the total variation 0 on any interval consisting of negative t -values. Thus $\tau = \omega$, as stated in (5).

17. If $\{\sigma_n\}$ is a given sequence of distribution functions, let σ_n be the convolution

$$\sigma_n = \bigstar_{k=1}^n \sigma_k = \sigma_1 * \sigma_2 * \cdots * \sigma_n,$$

and let the *infinite convolution* of the σ_n be defined formally as

$$\bigstar \sigma_n = \sigma_1 * \sigma_2 * \sigma_3 * \cdots.$$

The infinite convolution $\bigstar \sigma_n$ is said to be *convergent* if there exists a distribution function σ such that $\sigma_n \rightarrow \sigma$, in which case one writes $\sigma = \bigstar \sigma_n$. On denoting by $\sigma_{n,m}$, where $n < m$, the distribution functions

$$\sigma_{n,m} = \sigma_{n+1} * \sigma_{n+2} * \cdots * \sigma_m$$

and supposing that $\sigma = \bigstar \sigma_n$ is convergent, one sees that not only $\sigma_n \rightarrow \sigma$ as $n \rightarrow \infty$ but also $\sigma_n * \sigma_{n,m} = \sigma_m$, no matter how $m > n$ depends on n . Thus, by (5), $\sigma_{n,m} \rightarrow \omega$ as $n \rightarrow \infty$, for arbitrary $m = m_n$. This proves one-half of the following Cauchy criterion for convergence of $\bigstar \sigma_n$:

(6) *The infinite convolution $\bigstar \sigma_n$ is convergent if and only if $\sigma_{n,m} \rightarrow \omega$ as $n \rightarrow \infty$ no matter how $m > n$ depends on n ; cf. [10], Theorem 1.*

In order to prove the second half of (6), let it be assumed that the sequence σ_n is not convergent, so that, by (2), there exists an $a > 0$, such that, for every n , there is an $m > n$ for which $|\sigma_m(t') - \sigma_n(t'')| > a$ holds for every t', t'' in at least one t -interval of length $3a$. By the assumption that $\sigma_{n,m} \rightarrow \omega$, one may select this m and n in such a way that $|\sigma_{n,m}(+a) - \sigma_{n,m}(-a)| > 1 - a$. Now, if $t' - a, t', t' + a$ are continuity points of σ_n and σ_m in the above mentioned interval of length $3a$, one has

$$\sigma_{n,m}(t') = \int_{-\infty}^{-a} + \int_{+a}^{+\infty} + \int_{-a}^{+a} \sigma_n(t' - s) d\sigma_{n,m}(s),$$

where

$$0 \leq \int_{-\infty}^{-a} + \int_{+a}^{+\infty} \leq a \quad \text{and} \quad (1-a)\sigma_n(t' - a) \leq \int_{-a}^{+a} \leq \sigma_n(t' + a);$$

so that $(1-a)\sigma_n(t' - a) \leq \sigma_{n,m}(t') \leq \sigma_n(t' + a) + a$. A contradiction is

now obtained, since there must exist a t'' between $t' - a$ and $t' + a$ for which $|\sigma_m(t') - \sigma_n(t'')| \leq a$, while $|\sigma_m(t') - \sigma_n(t'')| > a$ for any such t'' .

If $\sigma = \star \sigma_n$ is convergent, it follows from (6) that, on placing $\sigma_n = \sigma_{n+1} * \sigma_{n+2} * \dots$, so that $\sigma = \sigma_n * \sigma_n$, one has $\sigma_n \rightarrow \omega$ as $n \rightarrow \infty$.

(6 bis) If $\star \sigma_n$ is convergent, then the (topological) limit of the spectrum of σ_n exists and is the spectrum of $\star \sigma_n$; cf. [10], Theorem 3.

This means that if for every $\epsilon > 0$, the interval $I_\epsilon : t_0 - \epsilon < t < t_0 + \epsilon$ contains points of the spectrum of σ_n for infinitely many n , then, for every ϵ , the interval I_ϵ contains points of the spectrum of σ_n for all but a finite number of n and t_0 is in the spectrum of σ . It is sufficient to prove that t_0 is in the spectrum of σ , since the other part of the statement follows then from $\sigma_n \rightarrow \sigma$. Thus it is sufficient to prove that, for every $\epsilon > 0$, one has $\sigma(t_0 + 2\epsilon) - \sigma(t_0 - 2\epsilon) > 0$.

Let n be chosen such that I_ϵ contains a point of the spectrum of σ_n i.e., such that $\sigma_n(t_0 + \epsilon) - \sigma_n(t_0 - \epsilon) > 0$, and let n be so large that $\sigma_n(\epsilon) - \sigma_n(-\epsilon) > 0$. The proof is now evident, since one obtains easily from $\sigma = \sigma_n * \sigma_n$ that

$$\begin{aligned} \sigma(t_0 + 2\epsilon) - \sigma(t_0 - 2\epsilon) \\ \geq [\sigma_n(t_0 + \epsilon) - \sigma_n(t_0 - \epsilon)][\sigma_n(+\epsilon) - \sigma_n(-\epsilon)] > 0. \end{aligned}$$

Thus, by means of (2) and (6 bis), the spectrum of $\sigma = \star \sigma_n$ may be determined from the spectra of the σ_n .

18. Let $f(x)$ be a μ -measurable function on a space X , which carries a measure μ such that $\mu X = 1$. Then the distribution function $\sigma(t)$ of $f(x)$ is defined by $f(t) = \mu A_t$, where A_t is the x -set $[f(x) < t]$. Clearly the discontinuity points of $\sigma(t)$ are the values of t at which $\mu B_t \neq 0$, where B_t is defined by $f(x) = t$. Moreover, $\sigma(t-0) = \mu A_t$ and $\sigma(t+0) = \mu A_t + \mu B_t$.

On the other hand, if a distribution function $\sigma(t)$ is given, then a function $f(x)$ may be defined on a suitable space X such that $\sigma(t)$ is the distribution function of $f(x)$. In fact, let X be the open interval $0 < x < 1$ and let μ be the Lebesgue measure on this interval. Then the "inverse function" of $\sigma(t)$, which may be defined on $0 < x < 1$ has σ as its distribution function.

If f is μ -measurable on X , and $\sigma(t)$ is its distribution function, one can express certain μ -integrals on X in terms of corresponding σ -integrals and conversely, as follows:

(7) One has

$$\int_X g(f(x)) dX = \int_{-\infty}^{+\infty} g(t) d\sigma(t),$$

if at least one of the integrals exists.

19. Suppose that $f_n(x_n)$ is μ_n -measurable on the space X_n , where $\mu_n X_n = 1$, and that the functions f_n are considered, according to I (7), as μ -measurable functions $f_n(x)$ on the product space $X = \Pi X_n$, where $\mu = \Pi \mu_n$. Let $\sigma_n(t)$ be the distribution function of $f_n(t)$.

The distribution function of $f_1 + f_2$ on X is $\sigma_1 * \sigma_2(t)$. In fact, if A_t is the X -set $[f_1(x) + f_2(x) < t]$, one has by I (8) and by the definitions of $\sigma_1(t), \sigma_2(t)$ as distribution functions of f_1, f_2 :

$$\mu(A_t) = \int_{f_1(x) + f_2(x) < t} dX = \int \int_{r+s < t} d\sigma_1(r) d\sigma_2(s) = \sigma_1 * \sigma_2(t - 0).$$

By an easy induction argument one obtains the proof of the following statement (if use is made of notations introduced in § 17):

(8) The distribution function of $f_1 + f_2 + \dots + f_n$ on X is σ_n and the distribution function of $f_{n+1} + f_{n+2} + \dots + f_m$ on X is $\sigma_{n.m}$.

The connection between Part II and the theory of infinite convolutions may be formulated as follows:

THEOREM V. The sequence Σf_n is almost everywhere convergent on $X = \Pi X_n$ if and only if the infinite convolution $\star \sigma_n$ of the distribution functions σ_n of the f_n is convergent, in which case $\sigma = \star \sigma_n$ is the distribution function of Σf_n ; cf. [10], Theorem 32.

In fact, by (6), $\star \sigma_n$ is convergent if and only if $\sigma_{n.m} \rightarrow \omega$ as $n \rightarrow \infty$, for arbitrary $m = m_n > n$. Now it is clear from (8) and from the definition of the distribution function of a function, that the condition $\sigma_{n.m} \rightarrow \omega$ is satisfied if and only if the sequence of functions $g_n(x) = f_n(x) + \dots + f_m(x)$ tends in measure to 0 on X , as $n \rightarrow \infty$, for arbitrary $m = m_n > n$. This is the case if and only if the series Σf_n is convergent in measure on X . Finally, by Theorem IV, Σf_n is convergent in measure on X if and only if Σf_n is convergent almost everywhere on X . The last statement of Theorem V is evident, since the definitions clearly imply that if s_n tends on X in measure to s , then $\tau_n \rightarrow \tau$, where τ_n, τ are the distribution functions of s_n, s respectively.

Another proof of Theorem V, which now follows, does not make use of the considerations of § 16 and § 17. In fact, it may be used to give a second proof of the statement (6) of § 17. One half of Theorem V is obvious. In fact, if Σf_n is convergent almost everywhere on X , then Σf_n is convergent in measure on X , so that $\{\sigma_n\}$ is a convergent sequence of distribution functions and $\sigma = \lim \sigma_n = \star \sigma_n$ is the distribution function of Σf_n . Now, suppose that $\star \sigma_n$ is convergent, so that $\{\sigma_{n.m}\}$ is a convergent sequence of distribution functions. If $\epsilon > 0$ is given, let $M = M_\epsilon$ be so large that $\sigma_n(M) - \sigma_n(-M) > 1 - \epsilon$

for every n . In the space $X = \Pi X_n$, let C^n be the set $C^n = [|s_n(x)| \leq M]$. Since, by (8), σ_n is the distribution function of $s_n = f_1 + \dots + f_n$, one has $\mu C^n \geq \sigma_n(M) - \sigma_n(-M) > 1 - \epsilon$. On selecting S_m in II (17) to be the partial sum s^n of Σf^n , one sees that $\Sigma(f_n - a_n)$ is convergent almost everywhere on X for a suitable choice of the a_n . Thus, by the half of Theorem V which has already been proved, $\star \sigma_n(t - a_n)$ is convergent. Since also $\star \sigma_n$ is convergent, the sequence Σa_n is convergent. And finally, Σf_n is convergent almost everywhere on X . This completes the second proof of Theorem V.

20. It is clear from (7) that if the k -th moment $N_k(\sigma_n)$ of σ_n is defined by

$$N_k(\sigma_n) = \int_{-\infty}^{\infty} t^k d\sigma_n(t),$$

then $M_k(f_n) = N_k(\sigma_n)$, if at least one of these moments exists. Moreover, if

$$\bar{N}_2(\sigma_n) = \int_{-\infty}^{+\infty} \{t - N_1(\sigma_n)\}^2 d\sigma_n(t) = N_2(\sigma_n) - N_1(\sigma_n)^2,$$

then $\bar{M}_2(f_n) = \bar{N}_2(\sigma_n)$. Similarly, if A_n denotes the x_n -set $[|f_n(x_n)| \leq K]$, then

$$\mu_n(X_n - A_n) = \sigma(-K - 0) + 1 - \sigma(K + 0);$$

and

$$\int_{A_n} g(f_n) dX_n = \int_{-K}^K g(t) d\sigma_n(t)$$

if at least one of the integrals exists.

As a consequence of Theorem V and the above remarks, one can write the convergence criteria of Part II for Σf_n as convergence criteria for $\star \sigma_n$. Accordingly, in the following list of theorems, the proofs are represented by references to statements in Part II or to preceding theorems in the same list.

THEOREM VI (Three series theorem). *The infinite convolution $\star \sigma_n$ is convergent if and only if, for a fixed $K > 0$, the following three series are convergent:*

$$\Sigma(1 + \sigma_n(-K) - \sigma_n(K)); \Sigma \int_{-K}^K t d\sigma_n(t); \Sigma \left\{ \int_{-K}^K t^2 d\sigma_n(t) - \left(\int_{-K}^K t d\sigma_n(t) \right)^2 \right\},$$

in which case the same holds for every $K > 0$. [Cf. Theorem I.]

The shortest proof of Theorem VI may apparently be given by noting first that the convergence of $\star \sigma_n$ is equivalent with the convergence in measure of Σf_n (using (8) and (6)), then applying II (2), (which obviously

retains its validity in the case of convergence in measure), thus reducing the proof of Theorem VI to the proof of (10), which may be obtained from the theory of Fourier-Stieltjes transforms; cf. [10], § 4 and Theorem 34.

Proofs of (9)-(11) may be obtained either from Theorem VI or from the corresponding statements in Part II and also from the theory of Fourier-Stieltjes transforms of distribution functions.

(9) If $\bar{N}_2(\sigma_n)$ exists for every n and if $\Sigma \bar{N}_2(\sigma_n)$ is convergent, then $\sigma = \star \sigma_n$ exists if and only if $\Sigma N_1(\sigma_n)$ is convergent, in which case $N_1(\sigma) = \Sigma N_1(\sigma_n)$ and $\bar{N}_2(\sigma) = \Sigma \bar{N}_2(\sigma_n)$. [Cf. II (4).]

(10) If $\sigma_n(-K) = 0$, $\sigma_n(K) = 1$ for every n , then $\sigma = \star \sigma_n$ exists if and only if both series $\Sigma N_1(\sigma_n)$ and $\Sigma \bar{N}_2(\sigma_n)$ are convergent. [Cf. II (6) and II (4).]

(11) The infinite convolution $\star \sigma_n$ is convergent if so are the series

$$\Sigma \int_{-\infty}^{+\infty} t d\sigma_n(t) \text{ and } \Sigma \int_{-\infty}^{+\infty} |t|^p d\sigma_n(t)$$

for a fixed p , $1 < p \leq 2$. [Cf. II (7).]

The necessary condition for the convergence of Σf_n in § 9 appears now in the following form

(12) If $q < p$, $c > 0$ and if for every n , $\sigma_n(t) = 1 - \sigma_n(-t)$ and

$$\int_{-\infty}^{+\infty} |t|^q d\sigma_n(t) \leq c \int_{-\infty}^{+\infty} |t|^p d\sigma_n(t),$$

then the convergence of $\star \sigma_n$ implies that

$$\Sigma \left(\int_{-\infty}^{+\infty} |t|^p d\sigma_n(x) \right)^{r/p} \quad (r = \text{Max}(q, 2))$$

is convergent. [Cf. II (12).]

21. The infinite convolution $\star \sigma_n$ is said to be absolutely convergent if it is convergent and remains convergent on arbitrary permutation of the σ_n . In view of Theorem V, this is the case if and only if Σf_n is almost everywhere convergent no matter in what fixed order the terms are taken. Thus one obtains Theorem VII (which may also be obtained as an immediate consequence of Theorem II), and (13).

THEOREM VII. The infinite convolution $\star \sigma_n$ is absolutely convergent if and only if, for a fixed $K > 0$, the three series

$$\Sigma(1 + \sigma_n(-K) - \sigma_n(K)); \quad \Sigma \left| \int_{-K}^K t d\sigma_n(t) \right|; \quad \Sigma \int_{-K}^K t^2 d\sigma_n(t)$$

are convergent, in which case the same holds for every $K > 0$, and the infinite convolution $\sigma = \star \sigma_n$ is independent of the order of the terms. [Cf. Theorem II.]

(13) The infinite convolution $\star \sigma_n$ is absolutely convergent if the two series

$$\Sigma \left| \int_{-\infty}^{+\infty} t d\sigma_n(t) \right| \quad \text{and} \quad \Sigma \int_{-\infty}^{+\infty} |t|^p d\sigma_n(t)$$

are convergent for a fixed p , $1 < p \leq 2$. [Cf. Theorem VII or II (7) and Theorem II.]

From the remark following Theorem II, one obtains the following statement:

(14) If $\star \sigma_n$ is convergent, then there exists a sequence $\{a_n\}$ of constants such that $\star \sigma_n(t - a_n)$ is absolutely convergent.

In the theory of infinite convolutions it is hard to distinguish between the two types of convergence discussed in § 7 and § 8. Thus the criteria corresponding to those of § 8 are listed here as criteria for absolute convergence of $\star \sigma_n$. For (16), cf. [34].

(15) The infinite convolution $\star \sigma_n$ is absolutely convergent if for a fixed $K > 0$ the two series

$$\Sigma(1 + \sigma_n(-K) - \sigma_n(K)) \quad \text{and} \quad \Sigma \int_{-K}^K |t| d\sigma_n(t)$$

are convergent, in which case the same holds for every $K > 0$. [Cf. Theorem III.]

(16) The infinite convolution $\star \sigma_n$ is absolutely convergent if the series

$$\Sigma \int_{-\infty}^{+\infty} |t|^p d\sigma_n(t)$$

is convergent for a fixed p , $0 < p \leq 1$. [Cf. II (11).]

22. Let \mathfrak{A} represent a class of Borel sets on the infinite t -axis which class is invariant under translations of the t -axis, and which includes, along with any sequence of sets A_n , the set ΣA_n .

Such classes are, for instance, the class \mathfrak{A}' of all enumerable sets; the class \mathfrak{A}'' of all Borel sets which are 0-sets in the ordinary sense; the class of all 0-sets according to any Hausdorff measure.

A distribution function $\sigma(t)$ will be called *pure* if it has, with reference to every class \mathfrak{M} , the following property: If the σ -measure of one set in an \mathfrak{M} is not 0, then the σ -measure of some set in this \mathfrak{M} is 1. If, for instance, σ is pure and has a discontinuity point, then the σ -measure of a set in \mathfrak{M}' is not 0, so that, according to the definition the σ -measure of some enumerable set is 1. Such a distribution function will be called purely discontinuous. It is easy to prove that σ is pure if the σ -measure of every set in \mathfrak{M}'' is 0, i. e., if σ is absolutely continuous. On the other hand, although a continuous, but not absolutely continuous, pure distribution function σ is always singular, a singular distribution function need not be pure. These notions allow the formulation of the following theorem.

THEOREM VIII (Pure Theorem). *If the σ_n are purely discontinuous and $\sigma = \star \sigma_n$ converges, then σ is a pure distribution function.*

Let f_n on X_n be, for every n , a function which has σ_n as distribution function. Since $\star \sigma_n$ is convergent, the series Σf_n is convergent almost everywhere on $X = \Pi X_n$. And since each σ_n is purely discontinuous, the function f_n may be assumed to attain an at most enumerable set of distinct values on X_n . Let M denote the (enumerable) modul of values of t generated by the values taken by all f_n , and if A is any t -set, let $M(+)A$ denote the set formed by the sums of any element in M and any element in A . It is clear that if A belongs to a class \mathfrak{M} , then so does $M(+)A$.

Now suppose that the set A of the class \mathfrak{M} is such that the σ -measure σA of A is positive, i. e. that $f = \Sigma f_n$ takes values in A on a set D of positive measure in $X = \Pi X_n$. If C denotes the subset of X , where f takes values in $A' = M(+)A$, then clearly C satisfies the requirements of I (6), so that the measure of C is either 0 or 1. Since $C \supset D$ and the measure of D is positive, this implies that the measure of C is 1. Finally, the measure of C in X is equal to the σ -measure of the t -set $A' = M(+)A$, so that $\sigma A' = 1$. This completes the proof that $\sigma = \star \sigma_n$ is a pure distribution function.

In view of the remarks at the beginning of this Section, Theorem VIII immediately implies the following statement:

(17) *If $\sigma = \star \sigma_n$ is a convergent infinite convolution of purely continuous distribution functions σ_n , then σ is either purely discontinuous or singular or absolutely continuous; cf. [10], Theorem 35.*

23. It seems to be very difficult to decide in general which of the cases of Theorem VIII take place for a given convergent infinite convolution $\star \sigma_n$ of purely discontinuous distribution functions σ_n . In other words, no general rule is known, to decide for a given class \mathfrak{M} whether or not the σ -measure of

each t -set in \mathfrak{A} is 0 or not. In this section a proof will be given of a theorem of P. Lévy, which gives the decision in case of the class \mathfrak{W}' of § 22. Note that, according to § 15, the expression σt denotes, for a distribution function σ and a real number t , the value of the σ -measure of t , i. e. the jump of σ at t .

The following lemma is related to the proof of Theorem VIII in [21] and to [15], Lemma 3.

(18) If $0 < d \leq 1$, $0 < 6\epsilon < d$, $l > 0$ and if the distribution functions λ, μ, ν have the properties

$$\lambda = \mu * \nu, \quad \lambda(a + 2l) - \lambda(a - 2l) < d + \epsilon, \quad \nu(l) - \nu(-l) > 1 - \epsilon,$$

while $\lambda a = d$ holds for some value a of t , then there exist real numbers b and $c = a - b$, such that

$$(i) \quad d - \epsilon < \mu b < \frac{d + \epsilon}{1 - \epsilon}; \quad (ii) \quad |c| < l; \quad (iii) \quad \nu c > 1 - \frac{6\epsilon}{d}.$$

If p, q , are the discontinuities of μ and ν , then one obtains from the definition of a convolution, since $\lambda = \mu * \nu$:

$$(*) \quad d = \lambda a = \sum_{p+q=a} (\mu p)(\nu q).$$

In (*) the terms for which $|a - p| < l$ (or $|q| < l$) satisfy $\sum \nu q \leq 1$, and the remaining terms satisfy $\sum \nu q < \epsilon$, $\sum \mu p \leq 1$. Hence

$$d < \epsilon + \max_{|a-p| < l} \mu p.$$

If this maximum is reached at b , one obtains (ii) and the left half of (i). Next, one has

$$\begin{aligned} d + \epsilon > \lambda(a + 2l) - \lambda(a - 2l) &\geq \int_{-l}^l [\mu(a - t + 2l) - \mu(a - t - 2l)] d\nu(t) \\ &\geq [\mu(a + l) - \mu(a - l)](1 - \epsilon), \end{aligned}$$

which implies the other half of (i) and in addition, using the left half of (i),

$$(1 - \epsilon)[\mu(a + l) - \mu b - \mu(a - l)] < 2\epsilon + \epsilon d - \epsilon^2 < 3\epsilon.$$

Separating now in (*) the terms where $|a - p| > l$ and the term $p = b$, $q = c$, from the other terms, one finds

$$d \leq \epsilon + \frac{3\epsilon}{1 - \epsilon} (1 - \nu c) + \frac{d + \epsilon}{1 - \epsilon} \nu c,$$

which clearly implies the remaining inequality (iii) of (18).

In the proof of Theorem IX, use will be made of the following statement (19) which is an immediate consequence of (18):

(19) If $\lambda = \mu_n * v_n$, for every n and $\mu_n \rightarrow \lambda$, so that $v_n \rightarrow \omega$ by (4), and if for some a , one has $\lambda a = d > 0$, then there exist real numbers b_n, c_n such that $b_n + c_n = a$, $b_n \rightarrow a$, $\mu_n b_n \rightarrow d$ and $c_n \rightarrow 0$, $v_n c_n \rightarrow 1$.

The characterization of the purely discontinuous case of Theorem VIII or (17) may be formulated as follows (cf. [21], Theorem XIII, in the proof below use has been made of a letter from B. Jessen; cf. [3 bis], footnote ¹⁸).

THEOREM IX. If $\sigma = \star \sigma_n$ is the convergent infinite convolution of the purely discontinuous distribution functions σ_n , then σ is purely discontinuous if and only if

$$(**) \quad \Pi d_n \neq 0, \text{ where } d_n = \text{Max}_t \sigma_n t.$$

The sufficiency of condition (**) for the existence of at least one discontinuity of σ is obvious, so that the corresponding statement of Theorem IX follows from Theorem VIII. It remains to prove that if σ has at least one discontinuity point, then (**) is satisfied.

Let $\sigma_n, \sigma_{n.}$ denote the same convolutions as in § 17. On applying (19) to the equation $\sigma = \sigma_n * \sigma_{n.}$, instead of to $\lambda = \mu_n * \lambda_n$, one obtains a sequence of numbers c_n , such that $c_n \rightarrow 0$, $\sigma_{n.} c_n \rightarrow 1$, as $n \rightarrow \infty$. Thus, replacing $\sigma_n(t)$ by $\sigma_n(t - c_{n-1} + c_n)$, one can suppose that $c_n = 0$ for every n , i. e., that $\sigma_{n.} 0 \rightarrow 1$, which clearly implies $\sigma_n 0 \rightarrow 1$. On omitting a finite number of terms, one may suppose that $\sigma_n 0 > \frac{1}{2}$ for every n , and $\sigma 0 > \frac{1}{2}$. Then:

$$\sigma_n 0 \geq d = \sigma 0 > \frac{1}{2}; \quad \sigma_n 0 \rightarrow 1; \quad \sigma_{n.} 0 \rightarrow \sigma 0; \quad \sigma_n 0 \rightarrow 1,$$

and also

$$\sum_{p \neq 0} \sigma_n p \leq 1 - d; \quad \sum_{p \neq 0} \sigma_n p \leq 1 - \sigma_n 0.$$

Thus, from (2),

$$\sigma_n 0 = \prod_{k=1}^n \sigma_k 0 + \sum_{k=2}^n \prod_{l=k+1}^n \sigma_l 0 \cdot \sum_{0 \neq p=-q} \sigma_{k-1} p \sigma_k q \leq \prod_{k=1}^n \sigma_k 0 + \sum_{k=2}^n \prod_{l=k+1}^n \sigma_l 0 \cdot (1 - d)(1 - \sigma_k 0),$$

so that

$$d = \sigma 0 \leq \sigma_n 0 \leq \prod_{k=1}^n \sigma_k 0 + (1 - d) \left(1 - \prod_{k=2}^n \sigma_k 0\right) < \prod_{k=1}^n \sigma_k 0 + 1 - d.$$

On letting n tend to infinity, one obtains $\Pi \sigma_n 0 \geq 2d - 1 > 0$, so that the proof of Theorem IX is complete.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

1. E. Borel, "Les probabilités dénombrables et leurs applications arithmétique," *Rendiconti del circolo matematico di Palermo*, vol. 27 (1909), pp. 247-271.
2. P. J. Daniell, "Integrals in an infinite number of dimensions," *Annals of Mathematics*, vol. 20 (1919), pp. 281-288.
3. A. Denjoy, "Sur les variables pondérées multipliables de M. Cantelli," *Comptes Rendus*, vol. 196 (1933), pp. 1712-1714.
- 3 bis. P. Erdős and A. Wintner, "Additive arithmetical functions and statistical independence," *American Journal of Mathematics*, vol. 61 (1939), pp. 713-721.
4. P. Hartman, E. R. van Kampen and Aurel Wintner, "Asymptotic distributions and statistical independence," *American Journal of Mathematics*, vol. 61 (1939), pp. 477-486.
5. E. K. Haviland, "On the inversion formula for Fourier-Stieltjes transforms in more than one dimension," I and II, *American Journal of Mathematics*, vol. 57 (1935), pp. 94-100 and pp. 382-388.
6. E. K. Haviland, "A note on a property of Fourier-Stieltjes transforms in more than one dimension," *American Journal of Mathematics*, vol. 57 (1935), pp. 567-572.
7. E. K. Haviland, "On the momentum problem for distribution functions in more than one dimension," I and II, *American Journal of Mathematics*, vol. 57 (1935), pp. 562-568 and vol. 58 (1936), pp. 164-168.
8. E. Hopf, "On causality, statistics and probability," *Journal of Mathematics and Physics*, vol. 13 (1937), pp. 51-102.
9. B. Jessen, "The theory of integration in a space of an infinite number of dimensions," *Acta Mathematica*, vol. 63 (1934), pp. 249-323.
10. B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.
11. M. Kac, "Sur les fonctions indépendantes," I, *Studia Mathematica*, vol. 6 (1936), pp. 46-58.
12. M. Kac and H. Steinhaus, "Sur les fonctions indépendantes," II, *Studia Mathematica*, vol. 6 (1936), pp. 59-66.
13. S. Kaczmarz and H. Steinhaus, "Le système orthogonal de M. Rademacher," *Studia Mathematica*, vol. 2 (1930), pp. 231-247.
14. E. R. van Kampen and A. Wintner, "On bounded convolutions," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 564-566.
15. E. R. van Kampen and A. Wintner, "On divergent infinite convolutions," *American Journal of Mathematics*, vol. 59 (1937), pp. 635-654.
16. K. Knopp, "Mengentheoretische Behandlung einiger Probleme der diophantischen Approximationen und der transfiniten Wahrscheinlichkeiten," *Mathematische Annalen*, vol. 95 (1925), pp. 409-426.
17. A. Khintchine and A. Kolmogoroff, "Ueber Konvergenz von Reihen deren Glieder durch den Zufall bestimmt werden," *Recueil de la Société Mathématique de Moscou*, vol. 32 (1925), pp. 668-677.
18. A. Kolmogoroff, "Ueber die Summen durch den Zufall bestimmter zufälliger Grössen," *Mathematische Annalen*, vol. 99 (1928), pp. 309-319, vol. 102 (1930), pp. 484-488.
19. A. Kolmogoroff, "Allgemeine Masstheorie und Wahrscheinlichkeitsrechnung," *Comptes Rendus de l'Académie Communiste* (1929), pp. 8-21.

20. A. Kolmogoroff, "Grundbegriffe der Wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik und ihre Grenzgebiete*, vol. 2, no. 3 (1933).
21. P. Lévy, "Sur les séries dont les termes sont des variables éventuelles indépendantes," *Studia Mathematica*, vol. 3 (1931), pp. 119-155.
22. Z. Łomnicki and S. Ulam, "Sur la theorie de la mesure dans les espaces combinatoires et son application au calcul des probabilité, I, Variables indépendantes," *Fundamenta Mathematica*, vol. 23 (1939), pp. 237-278.
23. J. Marcinkiewicz and A. Zygmund, "Sur les fonctions indépendantes," *Fundamenta Mathematica*, vol. 29 (1937), pp. 60-90.
24. J. Marcinkiewicz and A. Zygmund, "Quelques théoremes sur les fonctions indépendantes," *Studia Mathematica*, vol. 7 (1938), pp. 104-120.
25. R. E. A. C. Paley and A. Zygmund, "On some series of functions," *Proceedings of the Cambridge Philosophical Society*, vol. 26 (1930), pp. 337-357 and 458-474, vol. 28 (1932), pp. 190-205.
26. H. Rademacher, "Über die Verteilung gewisser konvergenzerzeugender Faktoren," *Mathematische Zeitschrift*, vol. 11 (1921), pp. 276-288 and "Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen," *Mathematische Annalen*, vol. 87 (1922), pp. 112-138.
27. F. Riesz, "Untersuchungen über Systeme integrierbarer Funktionen," *Mathematische Annalen*, vol. 69 (1910), pp. 449-497.
28. H. Steinhaus, "Les probabilités dénombrables et leur rapport à la théorie de la mesure," *Fundamenta Mathematica*, vol. 4 (1923), pp. 286-320.
29. H. Steinhaus, "Sur la probabilité de la convergence de séries," *Studia Mathematica*, vol. 2 (1930), pp. 21-39.
30. H. Steinhaus, "La théorie et les applications des fonctions indépendantes au sens stochastique," *Actualités Scientifiques et Industrielles*, Paris (1938).
31. C. Visser, "The law of nought-or-one in the theory of probability," *Studia Mathematica*, vol. 7 (1938), pp. 143-149.
32. N. Wiener, "The homogeneous chaos," *American Journal of Mathematics*, vol. 60 (1938), pp. 897-936.
33. A. Wintner, "Gaussian distribution functions and convergent infinite convolutions," *American Journal of Mathematics*, vol. 57 (1935), pp. 821-826.
34. A. Wintner, "A note on the convergence of infinite convolutions," *American Journal of Mathematics*, vol. 57 (1935), p. 839.
35. A. Wintner, "On a class of Fourier transforms," *American Journal of Mathematics*, vol. 58 (1936), pp. 15-20.
36. A. Wintner, "On the densities of infinite convolutions," *American Journal of Mathematics*, vol. 59 (1937), pp. 376-378.
37. A. Wintner, "On the smoothness of infinite convolutions of the type occurring in the theory of the Riemann zeta-function," *American Journal of Mathematics*, vol. 61 (1939), pp. 231-236.
38. A. Zygmund, "On the convergence of lacunary trigonometric series," *Fundamenta Mathematica*, vol. 16 (1930), pp. 90-107.

K-CYCLIC ELEMENTS.*

By J. W. T. YOUNGS.¹

The introduction, by G. T. Whyburn,² of the concept of cyclic element into the study of continuous curves proved so fruitful that it is only natural to hope for a further decomposition of a Peano space. In this connection Whyburn himself has defined cyclic elements of higher order where the space is a closed and bounded subset of Euclidean n -space.³ The approach is combinatorial. Recently an attack has been made on the problem from a purely point set theoretic standpoint by D. W. Hall.⁴ The approach taken in this note is similar to that of Hall in that the space is a cyclicly connected Peano space and the decomposition is point set theoretic.

To enlarge on this comment we shall survey, for a moment, the two equivalent notions of cyclic element from which the generalizations spring. In a Peano space a proper cyclic element M_p consists of the totality of points which are conjugate to a point p which is *not an end point or a cut point*. A proper cyclic element $M(a, b)$ may also be defined as the totality of points which are conjugate to each of two distinct points a and b which are conjugate to each other.⁵ The generalization of Hall may be said to be from the first definition while that here presented is from the second.

A large number of desirable properties are common to both concepts but it is also true that the generalizations do not seem to be all one might hope for. Sadly missing, for example, in any analogue to the theorem that the product of a cyclic element and a connected set is connected. On the other hand it is

* Received August 18, 1939; Revised January 8, 1940.

¹ The work on this paper was done while the author was in residence at Charlottesville. He wishes to take this opportunity to express his appreciation to Professor G. T. Whyburn and the Department of Mathematics at Virginia for their helpful cooperation.

² G. T. Whyburn, "Cyclicly connected continuous curves," *Proceedings of the National Academy of Science*, vol. 13 (1927), pp. 31-38; W. L. Ayres, "On the structure of a plane continuous curve," *Proceedings of the National Academy of Science*, vol. 13 (1927), pp. 650-657; C. Kuratowski and G. T. Whyburn, *Sur les éléments cycliques et leurs applications*, *Fundamenta Mathematicae*, vol. 16 (1930), pp. 305-331. The reader is asked to consult this last paper for the terminology of the subject and for an extensive bibliography.

³ G. T. Whyburn, "Cyclic elements of higher orders," *American Journal of Mathematics*, vol. 56 (1934), pp. 133-146.

⁴ The author has had the privilege of reading Hall's contribution in manuscript.

⁵ For a proof see Kuratowski and Whyburn, *loc. cit.*, p. 311.

hoped that this note may serve as an indication of a general direction of approach which seems to be adequate for certain purposes.

It will be noticed that the discussion is confined entirely to non-degenerate elements and the theorems proved are analogous to those true for non-degenerate cyclic elements in a Peano space. It follows that nothing is said about a "hyper-space," for the non-degenerate elements will not cover the space. These remarks, together with others in this note, give rise to problems which might be of some interest to the reader.

Bi-conjugacy.

1. 1. The space (which we shall denote by the symbol 1) is a cyclicly connected Peano space, or we may consider it as a non-degenerate cyclic element of a Peano space. In general, small letters will denote points of the space, while capitals will be reserved for sets of points.

1. 2. A point a is said to be bi-conjugate to a point b (notation: $a \sim b$) if for every pair of distinct points x_1 and x_2 different from a and b it is true that in $1 - (x_1 + x_2)$, $S_a = S_b$.⁶ In general, a point a is said to be k -conjugate to a point b (notation: $a \sim_k b$) if for every set of k distinct points x_1, \dots, x_k different from a and b it is true that in $1 - (x_1 + \dots + x_k)$, $S_a = S_b$.

1. 3. We shall confine the discussion almost exclusively to bi-conjugacy; nevertheless, a large number of the statements will be equally true for k -conjugacy. *Whenever a statement (with obvious modifications) is true for the more general concept we shall indicate this by prefacing the remark with an asterisk.* In every instance the proofs for the general statements are obvious modifications of those offered for bi-conjugacy.

1. 4. The first three of the usual axioms for an equivalence are obviously satisfied for bi-conjugacy. That is: (1) $a \sim b$ or $a \not\sim b$, (2) $a \sim a$, (3) $a \sim b$ implies $b \sim a$. The transitivity property is absent but we do have a modification of it.

*THEOREM. If $a \sim x_1 \sim \dots \sim x_n \sim b$ then any pair of points separating a from b must contain a point of the set x_1, \dots, x_n .⁷

Proof. Take any pair (p, q) distinct from a, x_1, \dots, x_n, b . Now

⁶ If we are considering $1 - (x_1 + \dots + x_k)$, then by S_a we shall mean the component of $1 - (x_1 + \dots + x_k)$ which contains a .

⁷ The author is indebted to W. L. Ayres for some valuable suggestions in connection with this theorem.

$a \sim x_1$, hence in $1 - (p + q)$ we have $S_a = S_{x_1}$. For the same reason, $S_{x_1} = S_{x_2}, \dots, S_{x_n} = S_b$. Hence $S_a = S_b$ and the pair (p, q) does not separate a from b .

(For k -conjugacy the theorem will read: If $a \sim x_1 \sim \dots \sim x_n \sim b$, then any set of k distinct points separating a from b must contain a point of the set x_1, \dots, x_n).

*COROLLARY. If $a \sim x_1 \sim \dots \sim x_l \sim b$, $a \sim y_1 \sim \dots \sim y_m \sim b$, and $a \sim z_1 \sim \dots \sim z_n \sim b$, where the x 's, y 's, and z 's constitute a set of $(l + m + n)$ distinct points; then $a \sim b$.⁸

Bi-cyclic elements.

2.1. The totality of points bi-conjugate to each of three distinct points which are bi-conjugate to each other constitutes a bi-cyclic element. That is, if a, b, c are distinct points, and $a \sim b$, $b \sim c$, $c \sim a$, then they generate a bi-cyclic element $M(a, b, c)$ and $x \in M(a, b, c)$ if and only if $x \sim a, b$ and c . (The totality of points k -conjugate to each of $(k + 1)$ distinct points which are k -conjugate to each other constitutes a k -cyclic element).

2.2. A bi-cyclic element need not be connected; in fact, it may consist of exactly three points. Let the space be the points on the circumference of a circle together with those on three cords which form an inscribed triangle. If the vertices of the triangle are a, b and c , then $M(a, b, c) = a + b + c$.

However, we do have the following

*2.3. THEOREM. Any point of a bi-cyclic element is bi-conjugate to any other point of it.

Proof. If $x, y \in M(a, b, c)$ then x and y are both bi-conjugate to a, b and c by 2.1, and so by 1.4 $x \sim y$.

This gives a degree of homogeneity indicated below.

*2.4. THEOREM. If the distinct points x, y, z are elements of $M(a, b, c)$, then $M(x, y, z) = M(a, b, c)$.

Proof. The notation $M(x, y, z)$ is justified by 2.3. If $p \in M(x, y, z)$, p is bi-conjugate to x, y, z , each of which is bi-conjugate to a . Therefore $p \sim a$, by 1.4. Similarly $p \sim b, c$. Therefore $p \in M(a, b, c)$. Hence $M(x, y, z) \subset M(a, b, c)$. Similarly $M(a, b, c) \subset M(x, y, z)$.

⁸ Throughout the paper a weaker form of this corollary will be used, the form in which $l = m = n = 1$. In a paper read at the April meeting of the American Mathematical Society in Chicago (1939), T. Radó called attention to the corresponding "weak transitivity" for conjugacy, thus the key to the situation here follows his remarks.

*COROLLARY. If $M(a, b, c) \cdot M(x, y, z)$ contains three distinct points (u, v, w) then $M(a, b, c) = M(x, y, z)$.

Both are identical to $M(u, v, w)$.

*2.5. THEOREM. If $M(a, b, c) \cdot M(x, y, z) = p + q$, then the pair (p, q) cuts the space.

Proof. Suppose the theorem false. Since $a \not\sim p \not\sim x$, $a \not\sim q \not\sim x$, and the pair (p, q) does not cut the space, $a \not\sim x$ by 1.4. Similarly, $a \not\sim y, z$. Therefore, $a \notin M(x, y, z)$. By the same argument $b, c \notin M(x, y, z)$ and by 2.4 $M(a, b, c) = M(x, y, z)$.

Sequences of bi-cyclic elements.

3.1. In connection with the point set theoretic properties of $M(a, b, c)$ we have seen that it may consist of exactly three points and so need not be connected. Another property is obtained from the following

*THEOREM. If $a_n \rightarrow a$, $b_n \rightarrow b$ and $a_n \not\sim b_n$, ($n = 1, 2, 3, \dots$), then $a \not\sim b$.

Proof. Consider any (p, q) distinct from a and b . Take two regions⁹ $U(a), U(b)$ without common points,¹⁰ both excluding p and q . There is an integer n such that $a_n \in U(a)$, $b_n \in U(b)$. Now in $1 - (p + q)$, $a_n \in S_a$, $b_n \in S_b$ and $S_{a_n} = S_{b_n}$. Therefore, $S_a = S_b$, and $a \not\sim b$.

*3.2. THEOREM. $M(a, b, c)$ is closed.

Proof. Suppose $p_n \rightarrow p$, $p_n \in M(a, b, c)$, ($n = 1, 2, 3, \dots$). We assert that $p \not\sim a$. The sequence a, a, a, \dots converges to a , $p_n \rightarrow p$ and $p_n \not\sim a$, ($n = 1, 2, 3, \dots$). By 3.1, $p \not\sim a$. Similarly $p \not\sim b$ and c . Therefore $p \notin M(a, b, c)$. Hence $M(a, b, c)$ is closed.

*3.3. LEMMA. If (1) U, V and W are three regions disjoint by pairs, (2) there are two sets of points (a, b, c) and (x, y, z) such that $a, x \in U$; $b, y \in V$; $c, z \in W$, (3) in each set, the points are bi-conjugate to each other; then it follows that any pair of the six points are bi-conjugate.

Proof. It will suffice to show that $x \not\sim b$. Take any (p, q) different from x and b . Some region U, V or W is entirely in $1 - (p + q)$. There is a point from each of the sets (a, b, c) and (x, y, z) in this region, hence S_x and S_a both intersect this region and so are identical.

⁹ A region is understood to be a connected open set. The notation $U(p)$ means a region containing p .

¹⁰ If there is no such pair of regions, then $a = b$ and so $a \not\sim b$.

*3.4. THEOREM. If $\{M(a_n, b_n, c_n)\}$ is a sequence of distinct bi-cyclic elements, then $\lim M(a_n, b_n, c_n)$ cannot consist of more than two distinct points.

Proof. If the theorem is false we may suppose that $a_n \rightarrow a$, $b_n \rightarrow b$, $c_n \rightarrow c$, where a , b and c are distinct points. Take disjoint regions $U(a)$, $U(b)$ and $U(c)$. There is an integer n_0 such that for $n > n_0$, $a_n \in U(a)$, $b_n \in U(b)$, $c_n \in U(c)$. Now $a \sim b$, $b \sim c$, $c \sim a$ by 3.1, and by 3.3 $a \sim b, c$. Therefore $a_n \in M(a, b, c)$. Similarly $b_n, c_n \in M(a, b, c)$. Therefore $M(a_n, b_n, c_n) = M(a, b, c)$; that is, the $M(a_n, b_n, c_n)$ are not all distinct.

Using this theorem we may assert something stronger.

*3.5. THEOREM. There are at most a denumerable number of bi-cyclic elements.

Proof. In each bi-cyclic element arbitrarily pick three distinct points. If for a certain bi-cyclic element the points are (a, b, c) then the element is $M(a, b, c)$. Now with each element $M(a, b, c)$ associate a positive real number $\gamma[M(a, b, c)] = \min [\rho(a, b), \rho(b, c), \rho(c, a)]$.¹¹ Take any $\delta > 0$. If there are an infinite number of bi-cyclic elements with associated γ greater than δ , then we may assume they are $M(a_n, b_n, c_n)$, $\gamma[M(a_n, b_n, c_n)] > \delta > 0$ and $a_n \rightarrow a$, $b_n \rightarrow b$, $c_n \rightarrow c$. Clearly a , b and c are distinct. But it is impossible for $\lim M(a_n, b_n, c_n)$ to consist of three points. The theorem is now obvious.

*COROLLARY. The points of $M(a, b, c)$ belonging to any other bi-cyclic element are denumerable.

No other bi-cyclic element can have more than two points in common with $M(a, b, c)$ by 2.4, and as there are only a denumerable number of bi-cyclic elements, the result follows.

Components of $1 - M(a, b, c)$.

4.1. Since each bi-cyclic element is a closed set the complement is open and so the components of the complement are open as the space is locally connected.

LEMMA. If $M(a, b, c)$ is a bi-cyclic element, and $U(a)$, $U(b)$, $U(c)$ are disjoint regions, then no component S of $1 - M(a, b, c)$ can have points in common with each of the three regions.

Proof. If the statement is false then $S + U(a) + U(b)$ is a region containing a and b , but not c . Hence there is an arc from a to b in it and the arc omits c .¹² Some point x on this arc is in S . In the region $S + U(c)$

¹¹ $\rho(a, b)$ is the distance from a to b .

¹² This is a well known fact first demonstrated by R. L. Moore, "Concerning continuous curves in the plane," *Mathematische Zeitschrift*, vol. 15 (1922), pp. 254-260.

consider an arc from x to c . There is a last point y on this arc which is also on the arc from a to b . Now $y \in S$ and there are three arcs connecting y to a , b and c , each pair of the arcs having only y in common. We assert that $y \not\sim a$. Take any pair (p, q) different from a and y . In $1 - (p + q)$, S_y contains at least one of the points a , b , c , and S_a contains the same one since a , b and c are bi-conjugate to each other. Hence $S_y = S_a$.

Similarly $y \not\sim b, c$. Therefore $y \notin M(a, b, c)$ which is contradictory to the fact that $y \in S$.

It is clear that this method of proof will not generalize even to tri-cyclic elements. As a matter of fact, for a cyclicly connected Peano space the corresponding result is not true for tri-cyclic elements.

4.2. THEOREM. *The frontier points of S are in $M(a, b, c)$ and there cannot be more than two of them.*

The proof is immediate from the lemma.

4.3. It is a well known fact that every region truly in a cyclicly connected Peano space has at least two frontier points, thus it is easy to see that

THEOREM. *The frontier of a component S of $1 - M(a, b, c)$ consists of precisely two points, and these two points (as a pair) cut the space.*

Remark. If the space 1 has the property that no set of $(k - 1)$ points cuts it, then a component S of the complement of a k -cyclic element $M(a_1, \dots, a_k)$ must have at least k frontier points. It might be conjectured that it will also have at most k frontier points, but no information seems available on this point.

4.4. THEOREM. *If M is a bi-cyclic element and S is a component of $1 - M$ having p and q for frontiers, then \bar{S} contains at most one bi-cyclic element containing p and q .*

Proof. If M_1 and M_2 are two bi-cyclic elements in \bar{S} containing p and q then since S is a region we have an arc in S from $m_1 \in M_1$ to $m_2 \in M_2$. It follows that $m_1 \sim m_2$ and so $M_1 = M_2$.

4.5. THEOREM. *If S is a component of $1 - M$ then any bi-cyclic element N is such that $N \subset \bar{S}$ or $N \cdot S = 0$.*

Proof. By 2.3 any point of a bi-cyclic element is bi-conjugate to any other point of it. Thus by 4.3 if $N \cdot S \neq 0$, $N \subset \bar{S}$.

4.6. THEOREM. *If $\{S_n\}$ is any sequence of components of the complement of a bi-cyclic element $M(a, b, c)$, then $d(S_n) \rightarrow 0$.*

Proof. If the theorem is false, then there is an infinite sequence of components with diameters greater than some positive δ . It is no restriction to assume that they are S_n , and (since they are connected) that there exist triples of points $x_n, y_n, z_n \in S_n$, ($n = 1, 2, 3, \dots$), such that $x_n \rightarrow x$, $y_n \rightarrow y$, $z_n \rightarrow z$, where x, y and z are distinct. It is clear that $x, y, z \in M(a, b, c)$. Take three disjoint regions $U(x), U(y), U(z)$. There exists an integer n such that $x_n \in U(x), y_n \in U(y), z_n \in U(z)$. That is, a component S_n of $1 - M(a, b, c)$ has points in common with each of the three regions. This is contrary to Lemma 4.1.

COROLLARY. *If $p \neq q$ there are only a finite number of bi-cyclic elements containing p and q .*

The proof follows from the above together with 4.4.

Continua of convergence.

***5.1. THEOREM.** *Every non-degenerate continuum of convergence is contained in some bi-cyclic element $M(a, b, c)$.*

Proof. If C is a continuum of convergence then no finite set of points in C separates C in 1. Certainly no two points outside C separate C . Hence there exist three points $a, b, c \in C$ which are bi-conjugate to each other. Every point of C is bi-conjugate to a, b and c , therefore $C \subset M(a, b, c)$.

5.2. THEOREM. *Suppose $\{K_n\}$ is a sequence of continua with the following properties: (1) $L = \lim K_n$ contains at least three points, (2) $L \cdot \sum_1^\infty K_n = 0$; then there exists a bi-cyclic element M such that $L = \lim [K_n \cdot M]$.*

Proof. Since we are dealing with a compact metric space any sequence of sets contains a convergent subsequence and from this fact it follows that there is a bi-cyclic element M which contains $\lim K_n$. As a matter of fact, the element in question is that of 5.1. The proof from this point on follows the pattern of the proof for a similar theorem concerning cyclic elements and so is omitted.¹³

The well known theorem alluded to above is used to show that the property of containing a continuum of convergence is cyclicly reducible. We cannot say that the property is bi-cyclicly reducible since in general the sets $[K_n \cdot M]$ will not be connected.

5.3. Since a bi-cyclic element may not be connected (in fact, a bi-cyclic element

¹³ See Kuratowski and Whyburn, *loc. cit.*, pp. 314-315.

can have a non-denumerable number of components each of the same diameter) we cannot hope for a generalization of the *cyclic connectivity theorem* within a bi-cyclic element, but a generalization is possible within the whole space. In fact, we have at our disposal in the literature the very tool adequate for this purpose. According to a result of Nöbeling,¹⁴ if two closed sets in a Peano space cannot be separated by the omission of any set of k points then there are $(k + 1)$ arcs connecting the sets, and the arcs are independent, except in that they might have the same end points when considered in pairs.

If we take x and y two points of a k -cyclic element, this means that no set of k points separates x from y in the space and so there are $(k + 1)$ independent arcs connecting x and y in the space.

Conclusion.

The reader will have noticed several unanswered questions by this time. We wish, in conclusion, to propose a further problem. One might arbitrarily call each single point of the space which is not in any bi-cyclic element a degenerate bi-cyclic element, or if a single point is in two or more bi-cyclic elements it too might be called a degenerate bi-cyclic element. In this fashion a space is covered by its bi-cyclic elements (degenerate and otherwise) so we can consider the hyper-space of bi-cyclic elements. One of the most beautiful results of the cyclic element theory is that the hyper-space is, in a sense, a dendrite. An analogous theorem for bi-cyclic elements would be extremely desirable. The fact that the frontier points of a component of the complement of a non-degenerate bi-cyclic element are two in number should play an important rôle in the attack.

THE OHIO STATE UNIVERSITY.

¹⁴ G. Nöbeling, "Eine Verschärfung des n -Beinsatzes," *Fundamenta Mathematicae*, vol. 18 (1932), pp. 23-38; N. E. Rutt, "Concerning the cut points of a continuous curve when the arc curve, AB , contains exactly N independent arcs," *American Journal of Mathematics*, vol. 51 (1929), pp. 217-246.

THE POLYTOPE 2_{21} , WHOSE TWENTY-SEVEN VERTICES CORRESPOND TO THE LINES ON THE GENERAL CUBIC SURFACE.*¹

By H. S. M. COXETER.

CONTENTS

	PAGE
1. Summary of properties of the 27 lines,	457
2. An alternative derivation of the cycles,	459
3. The representation by points in the affine plane,	460
4. The representation by points in real Euclidean six-space,	464
5. The representation by points in complex Euclidean three-space,	468
6. The 36 double-sixes and the polytope 1_{22} ,	470
7. Collineations which generate $[3^2, 2, 1]'$ according to a known abstract definition,	473
8. The representation by lines in $PG(3, 4)$,	476
9. The 120 trihedral pairs and the polytope 4_{21} ,	479

§§ 1-4 show how simple and natural is Schoute's representation of the 27 lines by the vertices of Gosset's six-dimensional polytope, and how easily various plane projections of the polytope can be drawn. In §§ 5 and 6 we find new coördinates for this polytope, and for Elte's related polytope 1_{22} . In § 7 we derive five quaternary collineations (Table III) which generate the simple group of order 25920 in a particularly elegant manner. § 8 connects these ideas with the lines on a special cubic surface in the finite geometry $PG(3, 4)$. Finally, in § 9, we show that the 240 vertices of Gosset's eight-dimensional polytope 4_{21} lie by sixes in forty planes which correspond to the "non-isotropic" planes of $PG(3, 4)$.

1. Summary of properties of the 27 lines. To construct the configuration of the lines on the general cubic surface, it is usual to begin with a double-six

$$\begin{array}{c} a_1 a_2 a_3 a_4 a_5 a_6 \\ b_1 b_2 b_3 b_4 b_5 b_6, \end{array}$$

where a_2, a_3, a_4, a_5, a_6 are five skew lines having a common transversal b_1 , and so on.² The remaining fifteen lines are $c_{12}, c_{13}, \dots, c_{56}$, where c_{12} is the line of intersection of the planes $a_1 b_2, a_2 b_1$. It is then found that c_{12} intersects c_{34} but is skew to c_{23} , and that the lines form thirty-five other double-sixes,

* Received September 15, 1939.

¹ Presented to the Society in two parts: April 16, 1938 and September 7, 1939.

² Schläfli, 22.

each of which could have been used just as well to build up the configuration. Every line intersects ten others, which form five intersecting pairs. The cubic surface therefore has 45 tritangent planes, each containing three of the lines.

The incidences of the lines are unchanged by the transposition of suffix numbers 1 and 2, which can be thought of as a re-naming of certain lines, namely the interchange of rows of the double-six

$$\begin{array}{cccccc} a_1 & b_1 & c_{23} & c_{24} & c_{25} & c_{26} \\ a_2 & b_2 & c_{13} & c_{14} & c_{15} & c_{16}. \end{array}$$

The transpositions (1 2), (2 3), (3 4), (4 5), (5 6), which we shall denote by P_1, P, O, N, N_1 , generate the symmetric group on the six suffix numbers. But this is not the whole group of automorphisms of the configuration. The operator,³ say Q , which interchanges rows of the double-six

$$\begin{array}{cccccc} c_{23} & c_{13} & c_{12} & a_4 & a_5 & a_6 \\ b_1 & b_2 & b_3 & c_{56} & c_{46} & c_{45} \end{array}$$

increases the order from $6!$ to $72 \cdot 6! = 51840$; in fact, it enables us to replace a_1, \dots, a_6 by either row of any one of the 36 double-sixes.

These six operators generate the group in a particularly simple form.⁴ Their product in any order is of period twelve;⁵ e. g., in the order N_1NOPP_1Q it is

$$\begin{aligned} R &= (1\ 2\ 3\ 4\ 5\ 6) \cdot Q \\ &= (a_1\ a_2\ a_3\ c_{56}\ c_{16}\ b_3\ b_4\ b_5\ b_6\ c_{23}\ c_{34}\ a_6) \\ &\quad \cdot (c_{45}\ a_4\ c_{46}\ c_{15}\ c_{26}\ b_2\ c_{12}\ b_1\ c_{13}\ c_{24}\ c_{35}\ a_5)(c_{14}\ c_{25}\ c_{36}). \end{aligned}$$

We note that R^4 permutes the 27 lines in nine cycles of three (each cycle belonging to a tritangent plane⁶), namely

$$\begin{aligned} R^4 &= (a_1\ c_{16}\ b_6)(a_2\ b_3\ c_{23})(a_3\ b_4\ c_{34})(c_{56}\ b_5\ a_6) \\ &\quad \cdot (c_{45}\ c_{26}\ c_{13})(a_4\ b_2\ c_{24})(c_{46}\ c_{12}\ c_{35})(c_{15}\ b_1\ a_5)(c_{14}\ c_{25}\ c_{36}). \end{aligned}$$

On the other hand, the operator $S \doteq (3\ 4)R^4(3\ 4)R$ permutes them in three cycles of nine. (We note that $S^3 = R^4$.) Another operator of the same kind is

$$\begin{aligned} S^{-2} &= (a_1\ a_2\ c_{56}\ c_{16}\ b_3\ b_5\ b_6\ c_{23}\ a_6) \\ &\quad \cdot (b_4\ c_{12}\ c_{24}\ c_{34}\ c_{35}\ a_4\ a_3\ c_{46}\ b_2) \\ &\quad \cdot (c_{14}\ b_1\ c_{13}\ c_{25}\ a_5\ c_{45}\ c_{36}\ c_{15}\ c_{26}), \end{aligned}$$

³ This is practically the "substitution T " of Burnside, 7, p. 301.

⁴ Coxeter, 10, pp. 163-165.

⁵ Coxeter, 11, p. 608.

⁶ The fact that the 27 lines are the complete intersection of the cubic surface with nine planes seems to have been first remarked by Baker, 1, p. 15.

in which the first cycle is the same as the first cycle of R , omitting the tritangent plane $a_3 b_4 c_{34}$.

2. An alternative derivation of the cycles. A more elementary (though somewhat artificial) procedure is based on the observation that any six lines which are the lines of two tritangent planes can be arranged as a cycle, each skew to its two neighbors but intersecting the remaining three. For instance, the planes $a_2 b_3 c_{23}$, $c_{56} b_5 a_6$ give the cycle

$$(a_2 c_{56} b_3 b_5 c_{23} a_6),$$

which is permuted by R^2 . Into this cycle, the three lines of any one of twelve

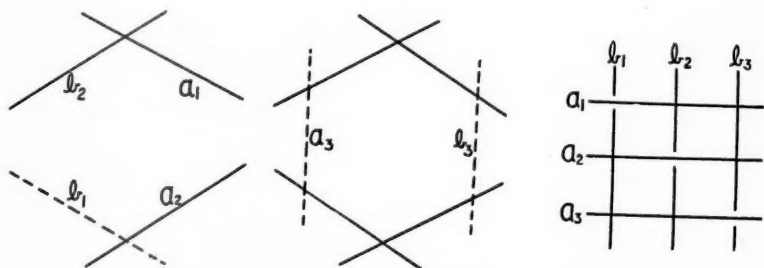


Fig. 1. The construction of a skew hexagon or double-three.

other tritangent planes can be inserted so that, in the consequent cycle of nine, each line is skew to its four neighbors (two before and two after) but intersects the remaining four (its four "opposites"). Inserting thus $a_1 c_{16} b_6$, we obtain the cycle

$$(a_1 a_2 c_{56} c_{16} b_3 b_5 b_6 c_{23} a_6),$$

which is permuted by S^{-2} .

Into this last cycle we can insert the three lines of any one of three tritangent planes (from among the eleven just discarded) so that, in the consequent cycle of twelve, each line is skew to its six neighbors (three before

and three after) but intersects its five opposites. Inserting thus $a_3 b_4 c_{34}$, we obtain the cycle

$$(a_1 a_2 a_3 c_{56} c_{16} b_3 b_4 b_5 b_6 c_{23} c_{34} a_6),$$

which is permuted by R , and whose square contains the original cycle of six lines.

3. The representation by points in the affine plane. Given two intersecting lines (from among the 27) and a third line skew to both, there is a

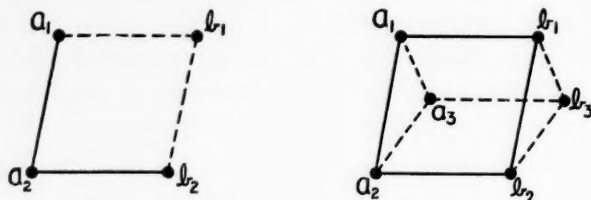


Fig. 2. The corresponding points.

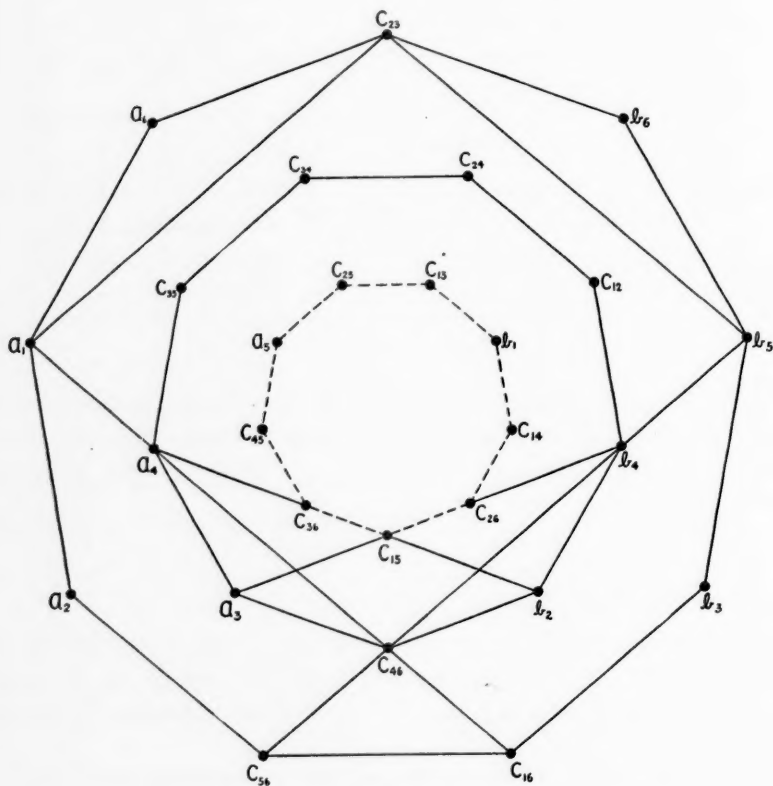


Fig. 3. Three enneagons.

uniquely determined fourth line, intersecting the third and again skew to the first two. For instance, the three lines a_1, b_2, a_2 determine b_1 , so as to form the intersecting pairs $a_1 b_2, a_2 b_1$ (Fig. 1). This suggests the possibility of

representing the 27 lines by 27 points, so that two such intersecting pairs are represented by the pairs of opposite vertices of a parallelogram, skew lines being represented by joined points. The transitivity of parallelism (Fig. 2) requires that any two other lines which intersect b_2, a_2 respectively, but not *vice versa*, should also intersect b_1, a_1 respectively, so as to form a skew hexagon or "double-three" (Fig. 1). This is in fact the case; for instance, the two

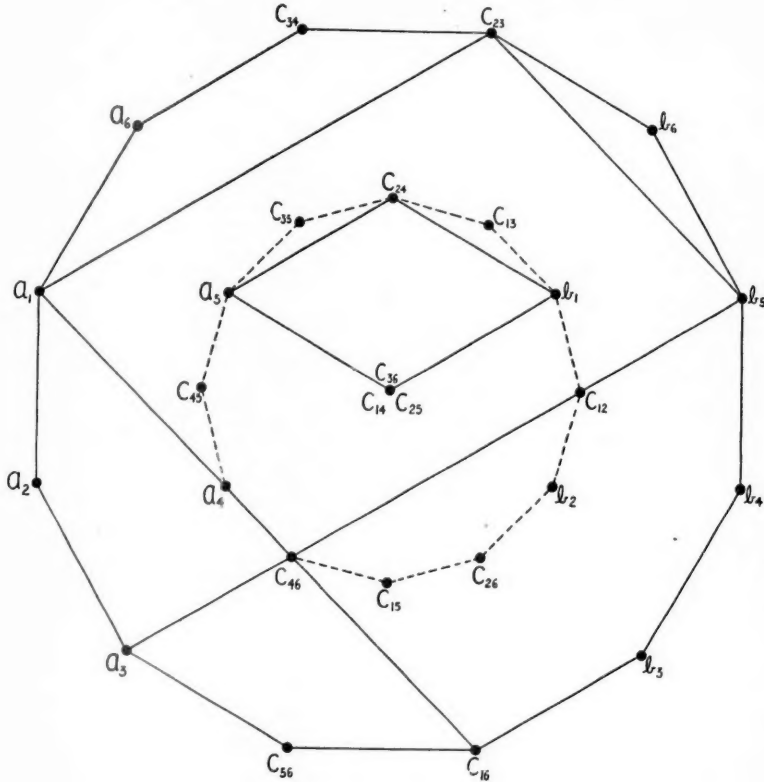


Fig. 4. Two dodecagons.

other lines may be a_3, b_3 , completing the skew hexagon $a_1 b_2 a_3 b_1 a_2 b_3$. Thus the representation is consistent.

In order to make a diagram of pleasing appearance in the Euclidean plane, we begin by drawing a regular enneagon or dodecagon, to represent the lines of a cycle of nine or twelve as described in § 2. The remaining representative points can then be derived by completing parallelograms. In Fig. 3, the three vertices a_1, c_{23}, b_5 of the outermost enneagon lead to c_{46} , which is thus seen to be the point of intersection of the joins $a_1 c_{16}, b_5 c_{56}$;

the rest of the smaller enneagon $c_{46} b_2 b_4 c_{12} c_{24} c_{34} c_{35} a_4 a_3$ can be obtained similarly, or can be marked at once by using the second cycle of the operator S^{-2} in § 1. Again, the three vertices a_3, c_{46}, b_2 of this second enneagon lead to c_{15} , which is thus seen to be the point of intersection of the joins $a_4 b_2, b_4 a_3$; the third enneagon $c_{15} c_{26} c_{14} b_1 c_{13} c_{25} a_5 c_{45} c_{36}$ corresponds to the last cycle

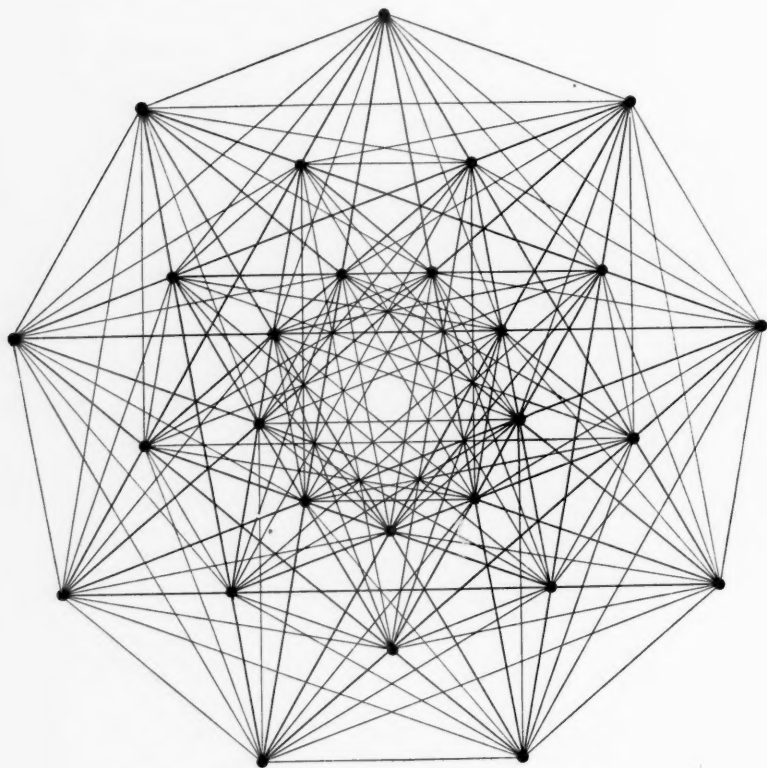


Fig. 5. The enneagonal projection of 2_{21} .

of S^{-2} . The complete diagram (Fig. 5, drawn by J. M. Andreas) has one unfortunate feature: 27 of the parallelograms (such as $c_{35} c_{25} c_{12} c_{13}$, $a_1 a_5 b_5 b_1$, $a_2 a_3 b_3 b_2$) degenerate into lines containing four points,⁷ thus we have to think of c_{25} as being joined to c_{12} but not to c_{13} , and so on.

In the dodecagon (Fig. 4), the three vertices a_1, c_{23}, b_5 lead to c_{46} , which is thus seen to be the point of intersection of the joins $a_1 c_{16}, b_5 a_3$; the smaller dodecagon $c_{46} c_{15} c_{26} b_2 c_{12} b_1 c_{13} c_{24} c_{35} a_5 c_{45} a_4$ can then be completed in ac-

⁷ This coincidence has been utilized by Rouse Ball, 2, p. 127.

cordance with the second cycle of R . Considerations of symmetry suffice to show that the remaining three points (corresponding to the last cycle of R) must coincide at the centre. In the complete diagram (Fig. 6), 24 of the parallelograms (such as $a_1 a_5 b_5 b_1$, $a_1 a_4 c_{16} c_{46}$) degenerate as before, while

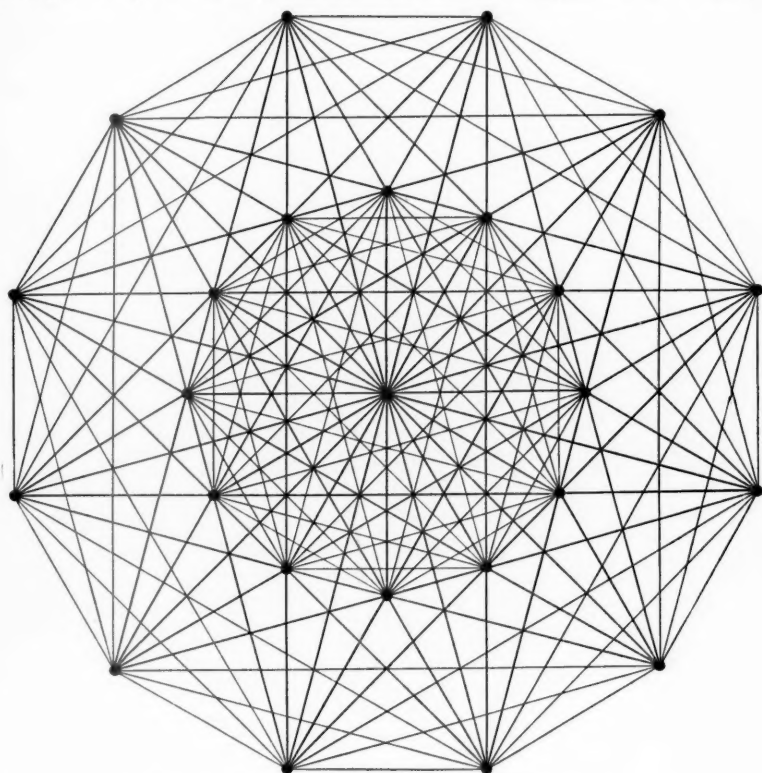


Fig. 6. The dodecagonal projection of 2_{21} .

another 24 (such as $a_1 c_{25} b_4 c_{36}$, $c_{24} c_{14} c_{15} c_{25}$) have two opposite vertices coincident (at the centre).

Figs. 7 and 8 both show the points that represent the ten lines meeting b_4 ,⁸ and those that represent the six skew lines b_i (i. e., one row of a double-six). Figs. 9 and 10 show the points that represent the nine lines

$$\begin{array}{l} c_{23} c_{15} c_{46} \\ a_2 b_1 c_{12} \\ b_3 a_5 c_{35} \end{array}$$

⁸ The particular line b_4 is chosen because $(a_1 a_2 a_3 c_{46} c_{14} c_{24} c_{34} a_6)$ is one cycle of the operator $(34)R$. We could obtain a third diagram by drawing this cycle as a regular octagon; then the three points b_4 , a_5 , c_{45} , would coincide at the centre.

of a trihedral pair.⁹ Fig. 9 makes it evident that three such sets of nine can exhaust the 27 lines, forming one of the forty *triads of trihedral pairs*.

4. The representation by points in real Euclidean six-space. Let us now make the representation more symmetrical by insisting that the parallelograms shall be squares. We can no longer remain in the plane; Fig. 2 has to

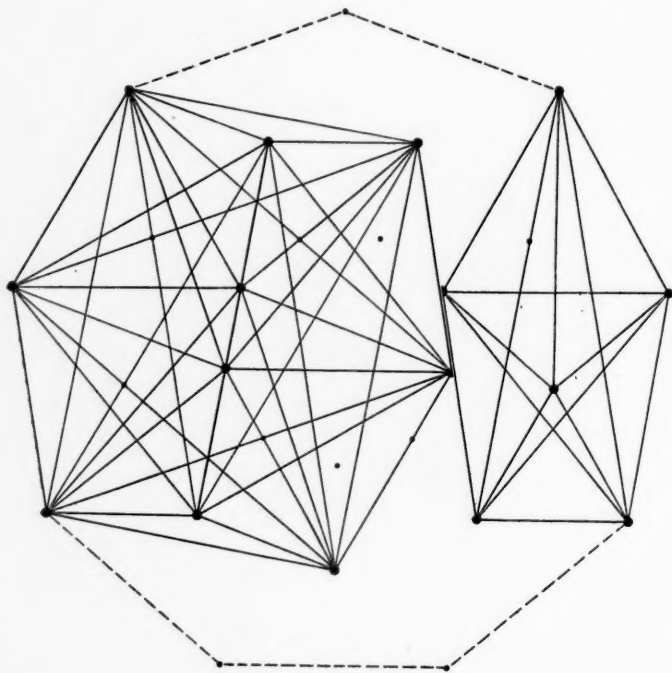


Fig. 7. β_5 and α_5 .

be regarded as a square and a triangular prism. The introduction of a_4 and b_4 necessitates a fourth dimension; of a_5 and b_5 , a fifth; of a_6 and b_6 , a sixth.

Let α_p denote the regular simplex (with $p + 1$ vertices) in p dimensions. Then the square and the triangular prism can be written as "rectangular products"¹⁰ $\alpha_1 \times \alpha_1$, $\alpha_2 \times \alpha_1$. The "double- n "

$$\begin{array}{c} a_1 a_2 \cdots a_n \\ b_1 b_2 \cdots b_n \end{array}$$

⁹ Steiner, 27.

¹⁰ Compare Coxeter, 11, pp. 591-592, where the rectangular product $\alpha_p \times \alpha_q$ is written $[\alpha_p, \alpha_q]$. Sommerville, (26, p. 114), calls this a "simplotope of type (p, q) ."

($n \leq 6$) is then represented by the rectangular product $\alpha_{n-1} \times \alpha_1$ in n dimensions, i. e., by a right prism whose base is the regular simplex α_{n-1} .

We are now representing the lines by points in six dimensions in such a way that the distance between two of the points is 1 or $2^{1/2}$ according as the corresponding lines are skew or intersect.¹¹ This applies to the c 's as well as to the a 's and b 's; for, we may take the representative points to have the following Cartesian coördinates in eight dimensions:¹²

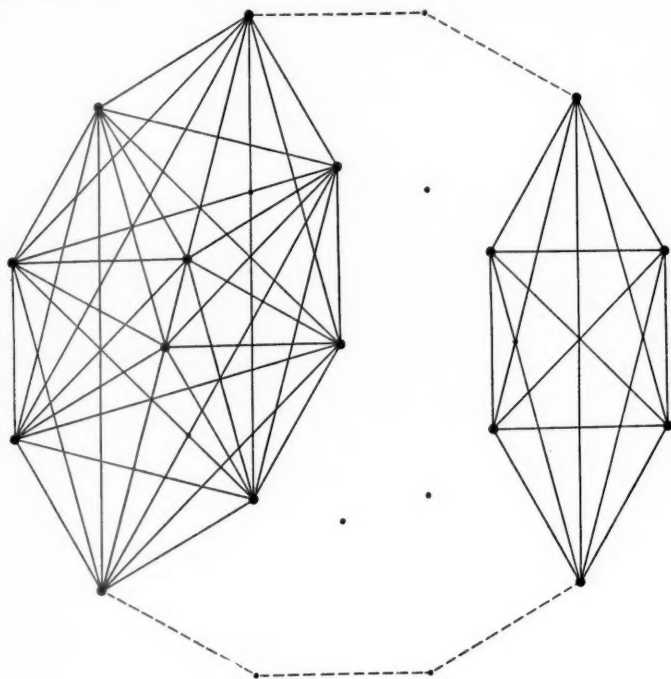


Fig. 8. β_5 and α_6 .

a_1	$(2k, 0, 0, 0, 0, 0, 2k, 0),$	b_1	$(2k, 0, 0, 0, 0, 0, 0, 2k),$	
a_2	$(0, 2k, 0, 0, 0, 0, 2k, 0),$	b_2	$(0, 2k, 0, 0, 0, 0, 0, 2k),$	
	$\cdot \cdot \cdot$		$\cdot \cdot \cdot$	
a_6	$(0, 0, 0, 0, 0, 2k, 2k, 0),$	b_6	$(0, 0, 0, 0, 0, 2k, 0, 2k),$	
c_{12}	$(-k, -k, k, k, k, k, k, k),$	$\cdot \cdot \cdot$	c_{56}	$(k, k, k, k, -k, -k, k, k).$

The number of dimensions is reduced from eight to six by the relations

$$x_1 + x_2 + \cdot \cdot \cdot + x_6 = x_7 + x_8 = 2k.$$

Since the distance $a_1 a_2$ is $2^{3/2}k$, we have $k = 2^{-3/2}$.

¹¹ To put it rather pedantically, the distance is $(r+1)^{1/2}$ when the two lines have r intersections. See Du Val, **15**, p. 28.

¹² Coxeter, **8**, pp. 3, 6.

This representation was discovered by P. H. Schoute and explained by J. A. Todd.¹³ It is perfect, in the sense that every automorphism of the 27 lines corresponds to a symmetry of the 27 points, and conversely.

We observe that all the above coördinates satisfy the condition $x_s \leq 2k$, equality holding for the five-dimensional simplex $b_1 b_2 b_3 b_4 b_5 b_6$ (corresponding to one row of a double-six); and that they all satisfy the condition

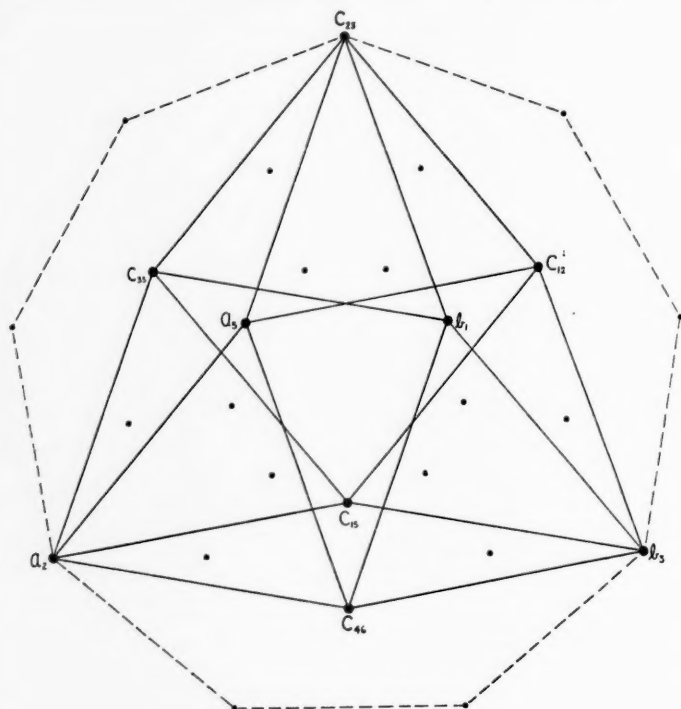


Fig. 9. $\alpha_2 \times \alpha_2$.

$x_1 + x_2 \leq 2k$, equality holding for the five-dimensional cross-polytope (or octahedron-analogue) whose pairs of opposite vertices are

$$a_1 b_2, a_2 b_1, c_{34} c_{56}, c_{35} c_{46}, c_{36} c_{45}$$

(corresponding to the ten lines which intersect c_{12}). Continuing thus, it can be shown that the 27 points in six dimensions are the vertices of a semi-regular polytope whose five-dimensional faces are regular polytopes of two kinds: 72 simplexes α_5 (belonging to the 36 inscribed $\alpha_5 \times \alpha_1$'s) and 27 cross-polytopes β_5 (one opposite to each vertex). The numbers of edges α_1 , triangles α_2 ,

¹³ Schoute, **24**, pp. 375-383; Todd, **28**.

tetrahedra α_3 , and "pentatopes" α_4 , are respectively 216, 720, 1080, and $432 + 216$.¹⁴

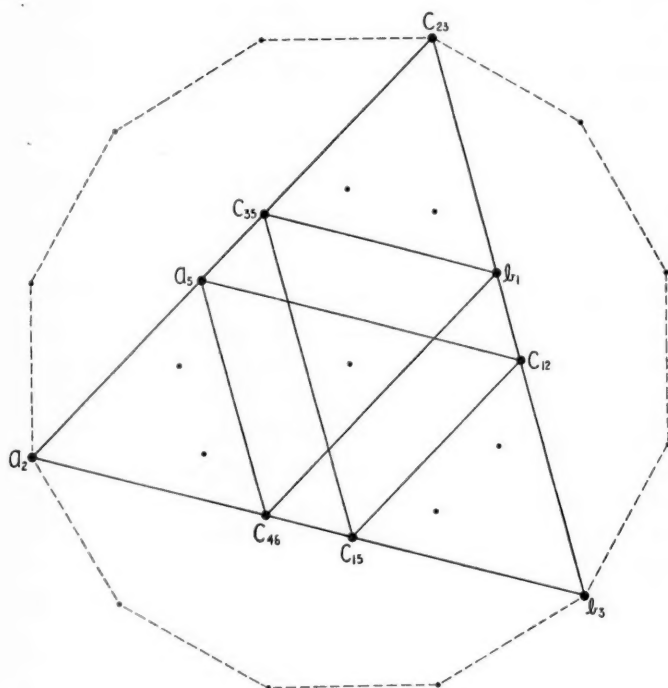
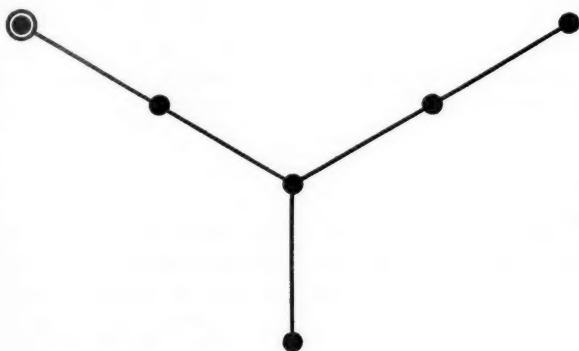


Fig. 10. $\alpha_2 \times \alpha_2$.

This six-dimensional polytope,¹⁵ now known as



¹⁴ Compare Henderson, **20**, p. 25.

¹⁵ Coxeter, **13**, p. 331. Cf. Fig. 11, below.

or 2_{21} , was discovered by Thorold Gosset.¹⁶ Figs. 5 and 6 can be regarded as plane projections of its vertices and edges. Squares such as $a_1 a_5 b_5 b_1$ are foreshortened into lines. Figs. 7 and 8 show two five-dimensional faces: a β_5 and an α_5 . The reader will have no difficulty in picking out from Figs. 5 and 6 (with the aid of Figs. 3 and 4, respectively) the prismatic figure $\alpha_5 \times \alpha_1$ whose vertices are all the a 's and b 's. Figs. 9 and 10 show an inscribed four-dimensional polytope, the rectangular product of two triangles, $\alpha_2 \times \alpha_2$, whose solid faces consist of six triangular prisms $\alpha_2 \times \alpha_1$ (or $\alpha_1 \times \alpha_2$), all plainly discernible. It is interesting to compare the different views in the two projections.

5. The representation by points in complex Euclidean three-space. Witting¹⁷ has shown that the simple group of order 25920 can be represented as a collineation group in four variables. Burkhardt¹⁸ used the transformations B, C, D, S_2 of Table I (at the end of this paper) to generate the corresponding linear group of order 51840, and remarked that the simple group itself is generated by linear transformations of the six Plücker coördinates, as in the middle section of the table. (The actual transformations are easily deduced from the first section by regarding $p_{\lambda\mu}$ as a symbolic product $z_\lambda z_\mu = -z_\mu z_\lambda$.¹⁹)

The final section of the table contains the corresponding transformations of

$$x_1 = p_{01} - \bar{p}_{23}, \quad x_2 = p_{02} - \bar{p}_{31}, \quad x_3 = p_{03} - \bar{p}_{12}.$$

These are not strictly "linear transformations," since D involves the conjugate imaginaries \bar{x}_2 etc.; however they are "unitary," in the sense of leaving $x_1 \bar{x}_1 + x_2 \bar{x}_2 + x_3 \bar{x}_3$ invariant. Consequently, if we write $x_\nu = y_\nu + y_{\nu+3}i$ where the y 's are real, the corresponding transformations of the six variables $y_1, y_2, y_3, y_4, y_5, y_6$ are orthogonal.²⁰ Using the terminology of the six-

¹⁶ Gosset gave this and many other new results in a long and brilliant essay which was refused publication in 1897. (For an abstract, see 19.) Since then, 2_{21} has been rediscovered at least three times.

¹⁷ 29.

¹⁸ 6, pp. 318, 320.

¹⁹ Dr. Frame has drawn my attention to the fact that this representation of degree six is an irreducible component of the Kronecker square of the representation by quaternary collineations, the other component being a representation of degree ten which is the corresponding symmetrized Kronecker square. It is also an irreducible component of the representation by permutations of the 27 lines on the cubic surface. He gives the character of the representation by quaternary collineations as $\pm 4, 0, 0, \pm 2, \pm (3\omega^2 + 1), \pm (3\omega + 1), \pm (\omega^2 - 1), \pm (\omega - 1), \pm \omega, \pm \omega^2, \pm 1, \pm (\omega - \omega^2), \pm 2, 0, 0, 0, 0, \pm 1, \pm \omega^2, \pm \omega$, taking the classes in the same order as in his table, 17, p. 483.

²⁰ Cf. Burkhardt, 6, p. 326.

dimensional Euclidean space defined by the y 's, we may say that the vector (X_1, X_2, X_3) is perpendicular to the hyperplane

$$u \equiv \bar{X}_1 x_1 + \bar{X}_2 x_2 + \bar{X}_3 x_3 + X_1 \bar{x}_1 + X_2 \bar{x}_2 + X_3 \bar{x}_3 = 0.$$

Moreover, if $X_1 \bar{X}_1 + X_2 \bar{X}_2 + X_3 \bar{X}_3 = 1$, the reflection in the hyperplane $u = 0$ is the transformation

$$(5.1) \quad \begin{cases} x'_1 = x_1 - X_1 u, \\ x'_2 = x_2 - X_2 u, \\ x'_3 = x_3 - X_3 u. \end{cases}$$

For, letting m, n take the values 1, 2, 3, 4, 5, 6, and writing $X_v = Y_v + Y_{v+3}i$ ($v = 1, 2, 3$), we have

$$\begin{aligned} 4 \Sigma Y_n y_n &= \Sigma (X_v + \bar{X}_v) (x_v + \bar{x}_v) - \Sigma (X_v - \bar{X}_v) (x_v - \bar{x}_v) \\ &= 2 \Sigma (\bar{X}_v x_v + X_v \bar{x}_v) = 2u; \end{aligned}$$

and the reflection in the hyperplane $\Sigma Y_n y_n = 0$ (where $\Sigma Y_n^2 = 1$) is

$$y'_m = y_m - 2 Y_m \Sigma Y_n y_n,$$

which leads at once to the above transformation (5.1).

Corresponding to the 27 lines on the cubic surface, Burkhardt²¹ found 27 linear complexes

$$(5.2) \quad \begin{cases} \bar{\omega}^\lambda p_{02} - \bar{\omega}^\mu p_{03} - \omega^\lambda p_{31} + \omega^\mu p_{12} = 0, \\ \bar{\omega}^\lambda p_{03} - \bar{\omega}^\mu p_{01} - \omega^\lambda p_{12} + \omega^\mu p_{23} = 0, \\ \bar{\omega}^\lambda p_{01} - \bar{\omega}^\mu p_{02} - \omega^\lambda p_{23} + \omega^\mu p_{31} = 0. \end{cases}$$

$$(\lambda, \mu = 0, 1, 2; \omega = e^{2\pi i/3}).$$

In terms of the x 's, these are the hyperplanes

$$\begin{aligned} \bar{\omega}^\lambda x_2 - \bar{\omega}^\mu x_3 + \omega^\lambda \bar{x}_2 - \omega^\mu \bar{x}_3 &= 0, \\ \bar{\omega}^\lambda x_3 - \bar{\omega}^\mu x_1 + \omega^\lambda \bar{x}_3 - \omega^\mu \bar{x}_1 &= 0, \\ \bar{\omega}^\lambda x_1 - \bar{\omega}^\mu x_2 + \omega^\lambda \bar{x}_1 - \omega^\mu \bar{x}_2 &= 0, \end{aligned}$$

which are perpendicular to the vectors

$$(5.3) \quad (0, \omega^\lambda, -\omega^\mu), \quad (-\omega^\mu, 0, \omega^\lambda), \quad (\omega^\lambda, -\omega^\mu, 0).$$

By keeping this selection of signs, we now have 27 points of the complex Euclidean (or "unitary") three-space which are permuted among themselves by the transformations B, C, D, S_2 of the x 's (see Table I).

In the real Euclidean six-space defined by the y 's we thus find 27 points, which correspond to the lines on the cubic surface, and which are permuted

²¹ 6, p. 323 (14).

among themselves by a rotation group of order 25920. We may naturally expect these to be *the vertices of the polytope* 2_{21} . Such is in fact the case. For, by considering various pairs of the points, we find that the distance²² between (x_1, x_2, x_3) and (x'_1, x'_2, x'_3) is $6\frac{1}{2}$ if $x_v + x'_v = 0$ for just one of the three values of v , and $3\frac{1}{2}$ otherwise. This agrees with the known correspondence between the 27 lines and the vertices of 2_{21} (edge $3\frac{1}{2}$) if we make the points (5.3) represent the lines

$$t_\lambda u_\mu, \quad u_\lambda s_\mu, \quad s_\lambda t_\mu$$

respectively, in the notation of Philip Hall.²³ The three lines of a tritangent plane (such as $t_0 u_0, u_0 s_0, s_0 t_0$, or $s_0 t_0, s_1 t_1, s_2 t_2$) are represented by the vertices of an equilateral triangle whose centre is the origin, i. e.²⁴ by three vectors whose sum is $(0, 0, 0)$.

6. The 36 double-sixes and the polytope 1_{22} . As we remarked in § 1, the whole group of automorphisms of the 27 lines, of order 51840,²⁵ can be generated by certain operators each of which interchanges the two rows of a double-six. Since the double-six is represented by the prismatic figure $\alpha_5 \times \alpha_1$ (§ 4), it is geometrically evident that the corresponding orthogonal transformations in Euclidean six-space are reflections which interchange pairs of opposite α_5 's of 2_{21} . These may equally well be described as reflections which interchange pairs of opposite *vertices* of the "semi-reciprocal" polytope 1_{22} .

This six-dimensional polytope, whose 72 vertices are the centres of the α_5 's of 2_{21} , was discovered by Elte.²⁶ The numbers of edges, triangles and tetrahedra are respectively 720, 2160 and 2160. Four-dimensional elements of two kinds are involved, namely 432 α_4 's and 270 β_4 's; but the five-dimensional faces are all alike, being 54 "half-measure-polytopes"²⁷ 1_{21} . From

$$^{22} (|x_1 - x'_1|^2 + |x_2 - x'_2|^2 + |x_3 - x'_3|^2)^{1/2}.$$

²³ See Coxeter, 9, p. 396. (It is obviously immaterial whether we take the suffix numbers to be 0, 1, 2 or 1, 2, 3). In this notation an operator of period nine (such as the S or S^{-2} of § 1) loses its artificiality, being expressible as $(s_0 t_0 u_0 s_1 t_1 u_1 s_2 t_2 u_2)$; thus the cycle of nine lines required in the construction of Fig. 5 is simply

$$(s_0 t_1 t_0 u_1 u_0 s_2 s_1 t_2 t_1 u_2 u_1 s_0 s_2 t_0 t_2 u_0 u_2 s_1).$$

²⁴ Cf. Frame, 18, p. 660. We shall have further comments to make on that paper in § 8.

²⁵ This must not be confused with the above mentioned linear group of order 51840, which has the simple group as a factor group but not as a subgroup; nor with Witting's collineation group of order 51840, which is a direct product, as was pointed out by Maschke, 21, p. 321.

²⁶ 16, pp. 104-108.

²⁷ The vertices of the $(n+3)$ -dimensional polytope 1_{n1} are *alternate* vertices of the

each of the 72 vertices emanate twenty edges which belong in pairs to ten regular hexagons lying in planes through the centre; hence there are altogether $72 \cdot 10/6 = 120$ such diagonal hexagons. The 27 pairs of opposite 1_{21} 's correspond to the lines on the cubic surface, the 36 pairs of opposite vertices correspond to the double-sixes, and the 120 diagonal hexagons correspond to the trihedral pairs.²⁸ Moreover, the planes of these hexagons fall into 40 sets of three absolutely perpendicular planes, corresponding to the triads of trihedral pairs.

New coördinates for the vertices of 1_{22} can of course be derived from those which we found for 2_{21} in § 5, but it is perhaps more interesting to obtain them from the collineation group. Corresponding to the 36 double-sixes, Burkhardt²⁹ found 36 linear complexes

$$(6.1) \quad \begin{cases} 3\frac{1}{2}i(\bar{\omega}^\lambda p_{01} + \omega^\lambda p_{23}) = 0, \\ 3\frac{1}{2}i(\bar{\omega}^\lambda p_{02} + \omega^\lambda p_{31}) = 0, \\ 3\frac{1}{2}i(\bar{\omega}^\lambda p_{03} + \omega^\lambda p_{12}) = 0, \\ \bar{\omega}^\kappa p_{01} + \bar{\omega}^\lambda p_{02} + \bar{\omega}^\mu p_{03} - \omega^\kappa p_{23} - \omega^\lambda p_{31} - \omega^\mu p_{12} = 0. \end{cases}$$

($\kappa, \lambda, \mu = 0, 1, 2$).

In terms of the x 's, these are the hyperplanes

$$(6.2) \quad \begin{cases} 3\frac{1}{2}i(\bar{\omega}^\lambda x_v - \omega^\lambda \bar{x}_v) = 0, \\ \bar{\omega}^\kappa x_1 + \bar{\omega}^\lambda x_2 + \bar{\omega}^\mu x_3 + \omega^\kappa \bar{x}_1 + \omega^\lambda \bar{x}_2 + \omega^\mu \bar{x}_3 = 0, \end{cases}$$

which are perpendicular to the vectors

$$(6.31) \quad (\pm 3\frac{1}{2}i\omega^\lambda, 0, 0), \quad (0, \pm 3\frac{1}{2}i\omega^\lambda, 0), \quad (0, 0, \pm 3\frac{1}{2}i\omega^\lambda),$$

$$(6.32) \quad (\pm \omega^\kappa, \pm \omega^\lambda, \pm \omega^\mu) \quad (\text{all } + \text{ or all } -).$$

In this case there is no systematic way of picking out one from each pair of oppositely directed vectors, but the 72 points having these coördinates are easily recognized as the vertices of 1_{22} (edge $3\frac{1}{2}$). In particular, the point $(1, 1, 1)$ corresponds to the simplex $(0, 1, -\omega^{\pm 1})$, $(-\omega^{\pm 1}, 0, 1)$, $(1, -\omega^{\pm 1}, 0)$, of 2_{21} .

Clearly, the six points $(\pm 3\frac{1}{2}i\omega^\lambda, 0, 0)$ are the vertices of a diagonal hexagon, and we have a simple verification of Todd's remark³⁰ that twelve of the 120 diagonal hexagons can be selected so as to include all the vertices of 1_{22} just once.

measure-polytope or hyper-cube, γ_{n+3} . Thus 1_{01} is the tetrahedron (having alternate vertices of the cube, γ_3), and 1_{11} is the cross-polytope β_4 ; but 1_{n1} is not regular for $n > 1$.

²⁸ Todd, 28, pp. 204-205.

²⁹ 6, p. 325.

³⁰ 28, p. 205.

To normalize the hyperplanes (6.2) we multiply by $3^{-1/2}$. The reflection in $i(\bar{\omega}^\lambda x_1 - \omega^\lambda \bar{x}_1) = 0$ is

$$(6.4) \quad \begin{cases} x'_1 = x_1 - (-i\omega^\lambda)i(\bar{\omega}^\lambda x_1 - \omega^\lambda \bar{x}_1) = \bar{\omega}^\lambda \bar{x}_1, \\ x'_2 = x_2, \\ x'_3 = x_3. \end{cases}$$

The reflection in $3^{-1/2}(x_1 + x_2 + x_3 + \bar{x}_1 + \bar{x}_2 + \bar{x}_3) = 0$ is

$$(6.5) \quad \begin{cases} x'_1 = x_1 - s, \\ x'_2 = x_2 - s, \\ x'_3 = x_3 - s, \end{cases}$$

where $s = (x_1 + x_2 + x_3 + \bar{x}_1 + \bar{x}_2 + \bar{x}_3)/3$.

The group of order 51840 which such reflections generate is conveniently denoted by $\begin{bmatrix} 3, 3 \\ 3, 3 \\ 3 \end{bmatrix}$ or (for brevity) $[3^2, 2, 1]$, since it has the abstract definition³¹

$$(6.6) \quad \begin{cases} O^2 = N^2 = N_1^2 = P^2 = P_1^2 = Q^2 \\ = (ON)^3 = (NN_1)^3 = (OP)^3 = (PP_1)^3 = (OQ)^3 \\ = (ON_1)^2 = (OP_1)^2 = (PQ)^2 = (QN)^2 = (NP)^2 \\ = (P_1Q)^2 = (QN_1)^2 = (N_1P)^2 = (NP_1)^2 = (N_1P_1)^2 = 1. \end{cases}$$

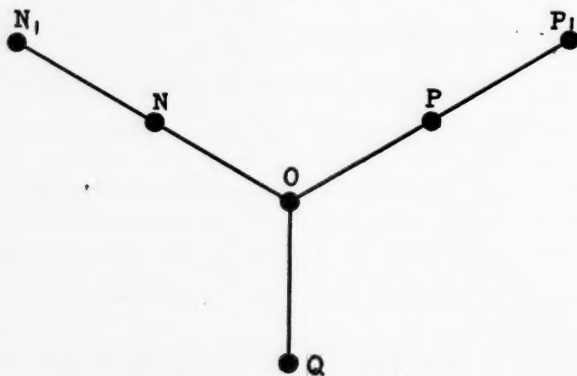


Fig. 11. The group $[3^2, 2, 1]$.

In order to generate the group in this elegant manner, we have to select six of the 36 reflecting hyperplanes in such a way that the angle between two of them is $\pi/2$ or $\pi/3$ according as the period of the product of the reflections is 2 or 3. In other words, we have to select six vertices of 1_{22} which are

³¹ Coxeter, 10, p. 164.

connected by five edges as in Fig. 11. In this diagram³² it is to be understood that pairs of vertices not joined by edges are distant $2\frac{1}{2}$ edge-lengths ($=6\frac{1}{2}$). N_1 denotes one of the two opposite vertices which are interchanged by the reflection N_1 ; similarly for the other letters.

Since the transformation (6.4) is simpler than (6.5), it is desirable to take as many as possible of the vertices from the set (6.31) and as few as possible from (6.32). The former set of vertices belong to three hexagons lying in absolutely perpendicular planes; so we cannot use them exclusively (since Fig. 11 is *connected*). We take NN_1 to be a side of one of these hexagons, PP_1 to be a side of another, and Q to be a vertex of the third. The remaining vertex, O , has to be chosen from the other set, and we naturally take it to be $(1, 1, 1)$. We thus obtain

$$\begin{array}{ll} N_1 & (-3\frac{1}{2}i, 0, 0), \\ N & (3\frac{1}{2}i\omega^2, 0, 0), \\ & O \quad (1, 1, 1), \\ & Q \quad (0, 0, 3\frac{1}{2}i\omega^2). \end{array} \quad \begin{array}{ll} P_1 & (0, -3\frac{1}{2}i, 0), \\ P & (0, 3\frac{1}{2}i\omega^2, 0), \end{array}$$

The corresponding reflections are given in the first section of Table II.

7. Collineations which generate $[3^{2,2,1}]'$ according to a known abstract definition. Writing $p_{01} - \bar{p}_{23}$, $p_{02} - \bar{p}_{31}$, $p_{03} - \bar{p}_{12}$ for x_1 , x_2 , x_3 , the transformation $x'_1 = x_1 - s$ of O (Table II) becomes

$$\begin{aligned} p'_{01} - \bar{p}'_{23} &= p_{01} - \bar{p}_{23} - (t + i), \quad \text{where} \\ t &= (p_{01} + p_{02} + p_{03} - p_{23} - p_{31} - p_{12})/3. \end{aligned}$$

This is consistent with either

$$p'_{01} = p_{01} - t, \quad p'_{23} = p_{23} + t$$

or

$$p'_{01} = -\bar{p}_{23} - i, \quad p'_{23} = -\bar{p}_{01} + i.$$

Thus the group $[3^{2,2,1}]$ is generated by collineations³³ in the six variables $p_{\lambda\mu}$, or equally well by *anticollineations*. (For details, see the second and third sections of Table II). In the case of the collineations it is impossible to regard the p 's as Plücker coördinates in a projective space defined by z_0, z_1, z_2, z_3 . But in the case of the anticollineations this can be done, as in the final section of the table. An ambiguity of sign appears at this stage, since reversing the signs of the z 's has no effect on the p 's. Hence although the group generated

³² The reader may be interested to draw plane projections of 1_{22} corresponding to the projection of 2_{21} shown in Figs. 5 and 6, and to pick out a set of five edges related as in Fig. 11.

³³ When working with the p 's there is no need to distinguish between collineations and linear transformations. (Burkhardt, 6, p. 320).

by anticollineations in the z 's is still $[3^{2,2,1}]$, the transformations themselves generate a group of order 103680 which we shall call the *binary* $[3^{2,2,1}]$.³⁴

The signs have actually been chosen so as to make the transformations satisfy (6.6) with the " $=1$ " omitted. These relations provide an abstract definition for the binary $[3^{2,2,1}]$. For, we could have continued them by writing " $=Z, Z^2=1$ "; but it is unnecessary to do so, since the relations $P^2=Q^2=(PQ)^2=Z$ imply $Z^2=1$ (and are fulfilled by the quaternion group).

The most important subgroups are $\{N_1, N, O, P, P_1\}$, of index 72; $\{N, O, P, P_1, Q\}$, of index 27; and $\{N_1N, N_1O, N_1P, N_1P_1, N_1Q\}$, of index 2. The last of these, having the abstract definition

$$(7.1) \quad \begin{cases} U^3 = V^2 = W^2 = X^2 = Y^2 \\ = (UV)^3 = (VW)^3 = (WX)^3 = (VY)^3 \\ = (UW)^2 = (UX)^2 = (UY)^2 = (VX)^2 = (WY)^2 = (XY)^2, \end{cases}$$

is naturally called the "binary $[3^{2,2,1}]$," since by adding the extra relation " $=1$ " we obtain $[3^{2,2,1}]'$, which is the simple group itself.³⁵ Table III (immediately deducible from Table II) gives linear transformations of the p 's which generate $[3^{2,2,1}]'$, and linear transformations of the z 's which generate the binary $[3^{2,2,1}]'$. These transformations, while possibly no more elegant than Burkhardt's, have the advantage of generating the groups according to a known abstract definition.

The operator of period twelve considered in § 1 is

$$(7.2) \quad R = UVWXY \\ = \begin{pmatrix} -\omega & 0 & 0 & 0 \\ 0 & -\omega & 0 & 0 \\ 0 & 0 & -\omega^2 & 0 \\ 0 & 0 & 0 & -\omega^2 \end{pmatrix} \begin{pmatrix} k & 0 & k & -k \\ 0 & k & k & k \\ k & k & -k & 0 \\ -k & k & 0 & -k \end{pmatrix}^{36} \begin{pmatrix} 0 & 0 & 0 & -\omega^2 \\ 0 & 0 & -\omega & 0 \\ 0 & \omega^2 & 0 & 0 \\ \omega & 0 & 0 & 0 \end{pmatrix} \\ \times \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \omega^2 & 0 \\ 0 & 0 & 0 & -\omega \\ -\omega & 0 & 0 & 0 \\ 0 & \omega^2 & 0 & 0 \end{pmatrix} \\ = \begin{pmatrix} -\omega k & -\omega k & \omega^2 k & 0 \\ -\omega k & \omega k & 0 & -k \\ \omega^2 k & 0 & k & -\omega k \\ 0 & -\omega^2 k & -k & -\omega k \end{pmatrix}.$$

³⁴ By analogy with the binary icosahedral group $P^5 = Q^2 = (PQ)^2$. (See Seifert and Threlfall, **25**, p. 218.) Since the binary icosahedral group arises as a linear group in two variables, and the binary $[3^{2,2,1}]$ as a linear group in four variables, it would perhaps have been better to name the latter the *quaternary* $[3^{2,2,1}]$.

³⁵ Coxeter, **10**, p. 160.

³⁶ $k = (\omega - \omega^2)/3 = 3^{-1/2}i$.

The central is generated by

$$Z = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

We have seen elsewhere³⁷ that the abstract group $[3^{2,2,1}]$ is generated by the two operators R and O . The results on which this statement is based apply to the binary $[3^{2,2,1}]$ without change. (For instance, $QOQ = O^{-1}Q^{-1}O^{-1}Z = OQO$, $QPQ = P^{-1}Z = P$.) It can be shown similarly that the binary $[3^{2,2,1}]'$ is generated by R and X . In fact, writing for brevity

$$\begin{aligned} T_n &= R^n X R^n, \text{ we have} \\ T_0 &= X = N_1 P_1, \\ T_5 &= R^{-5} N_1 R^5 \cdot R^{-5} P_1 R^5 = P_1 O, \\ T_2 &= O P O \cdot O Q O = O P Z Q O, \\ T_6 &= R^{-1} P_1 R \cdot R^{-1} O R = P N, \\ T_7 &= R^{-1} P R \cdot R^{-1} N R = O Q O N_1, \end{aligned}$$

whence

$$\begin{aligned} T_0 T_5 &= N_1 Z O, \\ T_7 T_0 T_5 &= O Q Z, \\ T_2 T_7 T_0 T_5 &= O P, \end{aligned}$$

and finally, since

$$Z = X^2 = X^{-2},$$

$$U = N_1 N = N_1 O \cdot O P \cdot P N = Z T_0 T_5 T_2 T_7 T_0 T_5 T_6 = X^{-1} R^{-5} X R^3 (X R^{-5})^3 X R^{-1} X R^6,$$

$$V = N_1 O = Z T_0 T_5 = X^{-1} R^{-5} X R^5,$$

$$W = N_1 P = N_1 Z O \cdot O P = T_0 T_5 T_2 T_7 T_0 T_5 = X R^{-5} X R^3 (X R^{-5})^3 X R^5,$$

$$Y = N_1 Q = N_1 O \cdot O Q Z = Z T_0 T_5 T_7 T_0 T_5 = X^{-1} R^{-5} X R^{-2} (X R^{-5})^2 X R^5.$$

In other words, the group of linear transformations, of order 51840, is generated by

$$\begin{pmatrix} -\omega k & -\omega k & \omega^2 k & 0 \\ -\omega k & \omega k & 0 & -k \\ \omega^2 k & 0 & k & -\omega k \\ 0 & -\omega^2 k & -k & -\omega k \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

By the same kind of argument, the second generator can be U instead of X .³⁸

Since the congruence

$$x^2 + x + 1 \equiv 0 \pmod{7}$$

³⁷ Coxeter, 11, p. 615.

³⁸ Brahana (4, p. 533) proved that the simple group $[3^{2,2,1}]'$ is generated by two operators which may be identified with our UVW and XY (or N_1NOP and P_1Q).

has the roots 2 and 4, one of the simplest modular representations³⁹ of the binary $[3^{2,2,1}]'$ is derived by putting 2 for ω , 4 for k ,⁴⁰ and regarding all the coefficients as residues modulo 7. Thus the group is generated by the transformations

$$U = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \quad V = \begin{pmatrix} 4 & 0 & 4 & 3 \\ 0 & 4 & 4 & 4 \\ 4 & 4 & 3 & 0 \\ 3 & 4 & 0 & 3 \end{pmatrix}, \quad W = \begin{pmatrix} 0 & 0 & 0 & 3 \\ 0 & 0 & 5 & 0 \\ 0 & 4 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{pmatrix},$$

$$X = \begin{pmatrix} 0 & 0 & 0 & 6 \\ 0 & 0 & 6 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \\ 5 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

in the field $GF[7]$, or alternatively by X (or U) and

$$R = \begin{pmatrix} 6 & 6 & 2 & 0 \\ 6 & 1 & 0 & 3 \\ 2 & 0 & 4 & 6 \\ 0 & 5 & 3 & 6 \end{pmatrix}.$$

The reader will probably agree that these matrices are easier to manipulate than those of (7.2).

8. The representation by lines in $PG(3, 4)$. Instead of $GF[7]$, we may use the field $GF[2^2]$ defined by the irreducible congruence

$$x^2 + x + 1 \equiv 0 \pmod{2}.$$

Tables I, II, III all remain valid if we interpret ω as a root of this congruence (a primitive root in the field) and define the conjugate \bar{u} of any mark u to be its square. Since $-1 \equiv 1$, we can replace every minus sign by plus. Also $k = (\omega - \omega^2)/3 \equiv \omega + \omega^2 \equiv 1$. Since $Z = 1$, there is no longer any distinction between collineations and linear transformations. We easily verify that the abstract definitions (6.6) and (7.1) (with " $= 1$ ") are satisfied. The transformations

$$U = \begin{pmatrix} \omega & 0 & 0 & 0 \\ 0 & \omega & 0 & 0 \\ 0 & 0 & \omega^2 & 0 \\ 0 & 0 & 0 & \omega^2 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}, \quad W = \begin{pmatrix} 0 & 0 & 0 & \omega^2 \\ 0 & 0 & \omega & 0 \\ 0 & \omega^2 & 0 & 0 \\ \omega & 0 & 0 & 0 \end{pmatrix},$$

³⁹ Brauer and Nesbitt, 5, p. 6.

⁴⁰ $k = (\omega - \omega^2)/3 \equiv (2 - 4)/3 \equiv 4 \pmod{7}$.

$$X = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 & \omega^2 & 0 \\ 0 & 0 & 0 & \omega \\ \omega & 0 & 0 & 0 \\ 0 & \omega^2 & 0 & 0 \end{pmatrix},$$

being unitary, generate $[3^{2,2,1}]'$ qua $HO(4, 2^2)$.⁴¹

Following Frame,⁴² let us regard these as collineations in the finite geometry $PG(3, 4)$ with homogeneous coördinates (z_0, z_1, z_2, z_3) . We observe that the line whose Plücker coördinates $(p_{01}, p_{02}, p_{03}, p_{23}, p_{31}, p_{12})$ are $(0, \omega^2, \omega^2, 0, \omega, \omega)$ is invariant under the transformations N_1, N, O, P, Q of Table II, and consequently also under the transformations U, V, W, Y of Table III (with the coefficients modified to fit the Galois field, as described above). This line is transformed into $(0, \omega, \omega^2, 0, \omega^2, \omega)$ by P_1 or by X , and into the 27 lines

$$(8.1) \quad \begin{cases} (0, \omega^\lambda, \omega^\mu, 0, \bar{\omega}^\lambda, \bar{\omega}^\mu), \\ (\omega^\mu, 0, \omega^\lambda, \bar{\omega}^\mu, 0, \bar{\omega}^\lambda), \\ (\omega^\lambda, \omega^\mu, 0, \bar{\omega}^\lambda, \bar{\omega}^\mu, 0), \end{cases}$$

by other operators of either $[3^{2,2,1}]$ or $[3^{2,2,1}]'$.

In other words, since the coefficients of the 27 linear complexes (5.2) satisfy

$$a_{01}a_{23} + a_{02}a_{31} + a_{03}a_{12} = -2 \equiv 0,$$

the corresponding linear complexes in $PG(3, 4)$ are special, each consisting of all lines which meet one of the lines (8.1). On the other hand, the 36 linear complexes (6.1) remain general, since for them the invariant is $-3 \equiv 1$.

Another consequence of reducing the coefficients to marks of $GF[2^2]$ is that the cubic form $z_0^3 + z_1^3 + z_2^3 + z_3^3$ is now invariant under the group $[3^{2,2,1}]'$ generated by U, V, W, X, Y ; it is transformed into its conjugate by the anticollineations which generate $[3^{2,2,1}]$. Thus the "cubic surface"

$$(8.2) \quad z_0^3 + z_1^3 + z_2^3 + z_3^3 \equiv 0$$

in the finite geometry $PG(3, 4)$ is invariant under a group of collineations and anticollineations which is simply isomorphic with the group of automorphisms of the lines on the general cubic surface in ordinary projective space. This result suggests the possibility that all the automorphisms of the lines on the special cubic surface (8.2) can be realized as collineations and

⁴¹ For alternative generators of $HO(4, 2^2)$, see Frame, 17, p. 482. For other representations of $[3^{2,2,1}]'$, see Dickson, 14, p. 298, and Brahana, *loc. cit.* (4, p. 533).

⁴² 18.

anticollineations. This is in fact the case, since the lines in question are precisely (8.1). The nine lines $(0, \omega^\lambda, \omega^\mu, 0, \bar{\omega}^\lambda, \bar{\omega}^\mu)$ belong to the trihedral pair⁴³ which is put in evidence by writing (8.2) in the form

$$(z_0 + z_1)(z_0 + \omega z_1)(z_0 + \omega^2 z_1) + (z_2 + z_3)(z_2 + \omega z_3)(z_2 + \omega^2 z_3) \equiv 0.$$

These "trihedra" are degenerate, each consisting of three planes through a line. Each of the planes is therefore met by the opposite "trihedron" in three concurrent lines (in contrast to a tritangent plane of the general cubic surface, in which the three lines form a triangle).

The same thing happens in the case of the special cubic surface

$$z_0^3 + z_1^3 + z_2^3 + z_3^3 = 0$$

in ordinary projective space. The 18 planes

$$z_\mu + \omega^\lambda z_\nu = 0 \quad (\mu < \nu)$$

each contain three concurrent lines of the surface. But the 27 planes

$$z_0 + \omega^\lambda z_1 + \omega^\lambda z_2 + \omega^\mu z_3 = 0$$

are proper tritangent planes, each containing a triangle. E.g. the plane $z_0 + z_1 + z_2 + z_3 = 0$ contains the triangle cut out on it by the planes

$$z_0 + z_1 = 0, \quad z_0 + z_2 = 0, \quad z_0 + z_3 = 0.$$

This dichotomy of the 45 tritangent planes indicates that the group of automorphisms of the lines on this special cubic surface is a proper subgroup of $[3^{2,2,1}]$. But the full symmetry is restored when we pass to the finite geometry by regarding the z 's as marks of $GF[2^2]$. For then all the 45 planes degenerate the same way; e.g. the planes

$$z_0 + z_1 + z_2 + z_3 \equiv 0, \quad z_0 + z_1 \equiv 0, \quad z_0 + z_2 \equiv 0, \quad z_0 + z_3 \equiv 0$$

concur at the point $(1, 1, 1, 1)$.

The configuration symbols for the whole space⁴⁴ $PG(3, 4)$ and for the cubic surface (8.2) are easily seen to be:

85	21	21	45	3	13
5	357	5	5	27	5
21	21	85	13	3	45

By this we mean that the "surface" contains 45 of the 85 points of the

⁴³ Cf. Frame, **18**, p. 661.

⁴⁴ Schoute, **23**, p. 5 ($m = 4$).

space, and 27 of the 357 lines; and that the 45 isotropic⁴⁵ planes of the space, each containing three of the 27 lines, may be regarded as "tangent" planes to the surface. The configuration thus determined is self-dual. There is a one-one correspondence between the 45 points and the 45 planes, each point being the point of concurrence⁴⁶ of the three lines in one of the planes. Each plane contains the corresponding point and twelve other points of the set. Each line contains five points and lies in the five corresponding planes.

The plane corresponding to (z_0, z_1, z_2, z_3) is (u_0, u_1, u_2, u_3) where $u_v = \bar{z}_v$. Hence the Plücker coördinates of the lines (after multiplication by ω or $\bar{\omega}$ if necessary) satisfy

$$\bar{p}_{01} = p_{23}, \quad \bar{p}_{02} = p_{31}, \quad \bar{p}_{03} = p_{12},$$

as in (8.1). Since $\bar{p}_{01}p_{23} = \bar{p}_{01}^2 = p_{01}$, the first three of the Plücker coördinates, so normalized, are the same as Frame's "non-homogeneous coördinates (4)," while the remaining three are their respective conjugates. Frame's rules for the incidences (Theorem 2) follow immediately. In particular, the three sides of a triangle on the general cubic surface correspond to two intersecting lines and a third line which is linearly dependent on them (i. e., concurrent and coplanar with them).

9. The 120 trihedral pairs and the polytope 4_{21} . Witting has shown⁴⁷ that the groups we have been considering have subgroups of index 40 of two distinct types. In the complex projective space with coördinates (z_0, z_1, z_2, z_3) , the plane $z_0 = 0$ is transformed into 40 planes, and the set of four planes $z_0 z_1 z_2 z_3 = 0$ is transformed into 40 tetrahedra. The vertices of the tetrahedra are 40 points whose coördinates are the same as the tangential coördinates of the planes. The 40 planes are⁴⁸

$$\begin{array}{ll} z_v = 0 & (v = 0, 1, 2, 3), \\ z_1 + \omega^\lambda z_2 + \omega^\mu z_3 = 0 & (\lambda, \mu = 0, 1, 2), \\ -z_0 - \omega^\mu z_2 + \omega^\lambda z_3 = 0, & \\ -\omega^\lambda z_0 + \omega^\mu z_1 - z_3 = 0, & \\ -\omega^\mu z_0 - \omega^\lambda z_1 + z_2 = 0, & \end{array}$$

or, when multiplied together, $F_{40} = 0$ in Maschke's notation.⁴⁹

When we interpret the coefficients as marks of the field $GF[2^2]$ (and

⁴⁵ Frame, **18**, p. 659.

⁴⁶ Thus the phrase "which form a triangle" should be deleted from the middle of p. 659 of Frame's paper.

⁴⁷ **29**, pp. 41-43. See also Burkhardt, **6**, p. 319.

⁴⁸ Blichfeldt, **3**, p. 151.

⁴⁹ **21**, p. 333.

therefore ignore the negative signs), these are precisely the 40 non-isotropic planes of $PG(3, 4)$. Hence, by Frame's Theorem 3,⁵⁰ the 40 tetrahedra correspond to the 40 triads of trihedral pairs of the cubic surface. In particular, the tetrahedron $z_0 z_1 z_2 z_3 = 0$ corresponds⁵¹ to the triad

$$\begin{array}{ccc|ccc|ccc} t_0 u_0 & t_1 u_2 & t_2 u_1 & u_0 s_0 & u_1 s_2 & u_2 s_1 & s_0 t_0 & s_1 t_2 & s_2 t_1 \\ t_1 u_1 & t_2 u_0 & t_0 u_2 & u_1 s_1 & u_2 s_0 & u_0 s_2 & s_1 t_1 & s_2 t_0 & s_0 t_2 \\ t_2 u_2 & t_0 u_1 & t_1 u_0 & u_2 s_2 & u_0 s_1 & u_1 s_0 & s_2 t_2 & s_0 t_1 & s_1 t_0 \end{array}$$

in Hall's notation.

When we interpret (z_0, z_1, z_2, z_3) as non-homogeneous coördinates (in complex Euclidean four-space), we find that the point $(3^{1/2}i, 0, 0, 0)$ is transformed into the 240 points

$$(9.1) \left\{ \begin{array}{l} (\pm 3^{1/2}i\omega^\lambda, 0, 0, 0), (0, \pm 3^{1/2}i\omega^\lambda, 0, 0), (0, 0, \pm 3^{1/2}i\omega^\lambda, 0), (0, 0, 0, \pm 3^{1/2}i\omega^\lambda), \\ (0, \pm \omega^\kappa, \pm \omega^\lambda, \pm \omega^\mu) \quad (\text{with signs agreeing}), \\ (\mp \omega^\kappa, 0, \mp \omega^\mu, \pm \omega^\lambda), \\ (\mp \omega^\lambda, \pm \omega^\mu, 0, \mp \omega^\kappa), \\ (\mp \omega^\mu, \mp \omega^\lambda, \pm \omega^\kappa, 0). \end{array} \right.$$

These correspond in sets of six to the 40 points considered above in the projective three-space, the six points of each set being derivable from any one of them by multiplying all four coördinates by the same power of $-\omega$. Thus $(\pm 3^{1/2}i\omega^\lambda, 0, 0, 0)$ is one such set of six, and $(0, \pm \omega^\lambda, \pm \omega^\lambda, \pm \omega^\lambda)$ is another.

When interpreted as points of real Euclidean eight-space, each set of six forms a regular hexagon. We thus have 40 hexagons lying in different planes through the origin. We shall see that the 240 points, so interpreted, have a far greater degree of symmetry than the 40 planes in which they lie; in other words, the 240 points can be distributed among 40 hexagons in many different ways. We shall prove, in fact, that the 240 points are the vertices of another of Gosset's semi-regular polytopes, namely



⁵⁰ Frame, 18, p. 660.

⁵¹ In detail, $(1, \omega^\lambda, 0, 0)$ and $(0, 0, 1, \omega^\mu)$ give the line $(0, 1, \omega^\mu, 0, \omega^{\lambda+\mu}, \omega^\lambda)$. To normalize this, we multiply by $\omega^{\lambda+\mu}$, obtaining $(0, \omega^{\lambda+\mu}, \omega^{\lambda-\mu}, 0, \omega^{-\lambda-\mu}, \omega^{-\lambda+\mu})$ or $(0, \omega^{\lambda+\mu}, \omega^{\lambda-\mu})$ or $t_{\lambda+\mu} u_{\lambda-\mu}$. To arrange nine such lines as a trihedral pair, we fix λ for the rows and μ for the columns.

or 4_{21} (of edge $3\frac{1}{2}$). This eight-dimensional polytope ⁵² has seven-dimensional faces of two kinds: 17280 simplexes α_7 , and 2160 cross-polytopes β_7 . The numbers of edges, triangles, tetrahedra, α_4 's, α_5 's, and α_6 's are respectively 6720, 60480, 241920, 483840, 483840, and $138240 + 69120$. Its symmetry group $[3^4, 2, 1]$, of order 696729600, is generated by reflections ⁵³ in hyperplanes which perpendicularly bisect the joins of pairs of opposite vertices.

Eight of these 120 reflections suffice, the abstract definition being ⁵⁴

$$\begin{aligned} O^2 &= N^2 = N_1^2 = N_2^2 = N_3^2 = P^2 = P_1^2 = Q^2 \\ &= (ON)^3 = (NN_1)^3 = (N_1N_2)^3 = (N_2N_3)^3 = (OP)^3 = (PP_1)^3 = (OQ)^3 \\ &= (ON_v)^2 = (OP_1)^2 = (PQ)^2 = (QN)^2 = (NP)^2 = (NN_2)^2 = (NN_3)^2 = (N_1N_3)^2 \\ &= (P_1Q)^2 = (QN_v)^2 = (N_vP)^2 = (NP_1)^2 = (N_vP_1)^2 = 1 \quad (v = 1, 2, 3), \end{aligned}$$

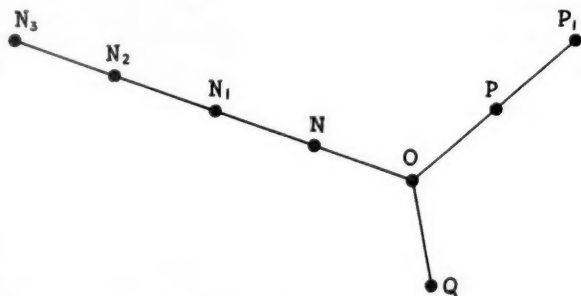


Fig. 12. The group $[3^4, 2, 1]$.

analogous to (6.6). There is a central of order two, whose quotient group is derived by inserting the extra relation

$$(N_3 N_2 N_1 N O P P_1 Q)^{15} = 1.$$

Seven rotations, such as

$$N_3 N_2, N_3 N_1, N_3 N, N_3 O, N_3 P, N_3 P_1, N_3 Q,$$

generate a subgroup of index two. This subgroup of the central quotient group ⁵⁵ of $[3^4, 2, 1]$ is the simple group $FH(8, 2)$, of order 174182400.

⁵² Gosset, 19, p. 48.

⁵³ Coxeter, 9, p. 388.

⁵⁴ Coxeter, 10, p. 171. The subgroup $[3^2, 2, 1]$ generated by O, N, N_1, P, P_1, Q has no direct connection with the binary $[3^2, 2, 1]$ generated by linear transformations of the z 's. (The O, N , etc. of Table II are of period four).

⁵⁵ Coxeter, 10, p. 174. There, and on p. 179, the word *sub-group* has several times been used by mistake for *factor group*.

In order to prove that the points (9.1) are the vertices of 4_{21} , we select eight of them whose mutual distances are indicated in Fig. 12. By comparison with Fig. 11, we easily find one such set of eight points to be

$$\begin{array}{ll} N_3 & (3\frac{1}{2}i, 0, 0, 0), \\ N_2 & (-\omega^2, \omega^2, 0, -\omega^2), \\ N_1 & (0, -3\frac{1}{2}i, 0, 0), \\ N & (0, 3\frac{1}{2}i\omega^2, 0, 0), \\ & O \quad (0, 1, 1, 1), \\ & Q \quad (0, 0, 0, 3\frac{1}{2}i\omega^2). \end{array} \quad \begin{array}{l} P_1 \quad (0, 0, -3\frac{1}{2}i, 0), \\ P \quad (0, 0, 3\frac{1}{2}i\omega^2, 0), \end{array}$$

The corresponding reflections are given in the first section of Table IV. It only remains to be observed that the given set of 240 points is invariant under these transformations.

When we pass to the real Euclidean eight-space, a more natural form of coördinates for Fig. 12 (on a different scale) is ⁵⁶

$$\begin{array}{ll} N_3 & (2, -2, 0, 0, 0, 0, 0, 0), \\ N_2 & (0, -2, 2, 0, 0, 0, 0, 0), \\ N_1 & (0, 0, 2, -2, 0, 0, 0, 0), \\ N & (0, 0, 0, -2, 2, 0, 0, 0), \\ & O \quad (0, 0, 0, 0, 2, 2, 0, 0), \\ & Q \quad (1, 1, 1, 1, 1, 1, 1, 1). \end{array} \quad \begin{array}{l} P_1 \quad (0, 0, 0, 0, 0, 0, -2, 2), \\ P \quad (0, 0, 0, 0, 0, 2, -2, 0), \end{array}$$

The group $[3^{4,2,1}]$ is then generated by reflections in the hyperplanes

$$\begin{array}{ll} y_0 = y_1, & \\ y_1 = y_2, & \\ y_2 = y_3, & y_6 = y_7, \\ y_3 = y_4, & y_5 = y_6, \\ y_4 + y_5 = 0, & \\ \Sigma y = 0. & \end{array}$$

(See the second section of Table IV).

Since the symmetry group of 4_{21} is $[3^{4,2,1}]$, of order 696729600, while that of the 40 hexagons is the binary $[3^{2,2,1}]$, of order 103680, it follows that such a set of 40 hexagons, which together use up all the 240 vertices of 4_{21} , can be selected in 6720 ways. Having made one such selection, we find that the planes of the 40 hexagons (which correspond to the non-isotropic planes of the finite geometry ⁵⁷) fall into 40 sets of four absolutely perpendicular

⁵⁶ For the consequent coördinates of all the vertices of 4_{21} (edge $2^{3/2}$), see Coxeter, 8, p. 2. ((PA)_s is an alternative symbol for 4_{21}).

⁵⁷ Frame, 18, p. 660 (§ 4).

planes. Hence, finally, *these forty sets of four planes correspond to the forty triads of trihedral pairs of the cubic surface.*

Another subgroup of index 6720 in $[3^4, 2, 1]$ can be derived by specializing a single one of the 1120⁵⁸ diagonal hexagons of 4_{21} (instead of forty of them). For, in (9.1), the plane of the hexagon $(\pm 3^{1/2}i\omega^\lambda, 0, 0, 0)$ is absolutely perpendicular to the six-space $z_0 = 0$, which contains the 72 points

$$(0, \pm 3^{1/2}i\omega^\lambda, 0, 0), \quad (0, 0, \pm 3^{1/2}i\omega^\lambda, 0), \quad (0, 0, 0, \pm 3^{1/2}i\omega^\lambda), \\ (0, \omega^\kappa, \omega^\lambda, \omega^\mu), \quad (0, -\omega^\kappa, -\omega^\lambda, -\omega^\mu).$$

By (6.3), these are the vertices of the six-dimensional polytope⁵⁹ 1_{22} , whose symmetry group, $[3^{2, 2, 1}] \times G_2$, is derived from that of 2_{21} by adjoining the central inversion. In the notation of Table IV, this subgroup is $\{N_3, N_1, N, O, P, P_1, Q\}$.

The 40 planes and 40 absolutely perpendicular six-spaces give, by intersection with a six-space of general position, a configuration of 40 points and 40 four-spaces, which is the real counterpart of Witting's configuration of 40 points and 40 planes in complex projective three-space. Witting described his configuration as the three-dimensional analogue of the Hessian configuration of $9 + 12$ points and $9 + 12$ lines in the complex projective plane. The real counterpart of the latter comes from the planes of nine diagonal triangles of 2_{21} and of twelve diagonal hexagons of the semi-reciprocal 1_{22} , together with the absolutely perpendicular four-spaces, by intersection with a four-space of general position.

TABLE I

The simple group $[3^{2, 2, 1}]$, of order 25920, as a collineation group (after Burkhardt)

	<i>B</i>	<i>C</i>	<i>D</i>	<i>S</i> ₂
$z'_0 =$	z_0	z_0	$-z_1$	$\omega^2 z_0$
$z'_1 =$	$-k(z_1 + z_2 + z_3)^*$	z_3	$-z_0$	z_1
$z'_2 =$	$-k(z_1 + \omega z_2 + \omega^2 z_3)$	z_1	$-z_2$	$\omega^2 z_2$
$z'_3 =$	$-k(z_1 + \omega^2 z_2 + \omega z_3)$	z_2	z_3	$\omega^2 z_3$
$p'_{01} =$	$-k(p_{01} + p_{02} + p_{03})$	p_{03}	$-p_{01}$	$\omega^2 p_{01}$
$p'_{02} =$	$-k(p_{01} + \omega p_{02} + \omega^2 p_{03})$	p_{01}	p_{12}	ωp_{02}
$p'_{03} =$	$-k(p_{01} + \omega^2 p_{02} + \omega p_{03})$	p_{02}	p_{31}	ωp_{03}
$p'_{23} =$	$k(p_{23} + p_{31} + p_{12})$	p_{12}	$-p_{23}$	ωp_{23}
$p'_{31} =$	$k(p_{23} + \omega^2 p_{31} + \omega p_{12})$	p_{23}	p_{03}	$\omega^2 p_{31}$
$p'_{12} =$	$k(p_{23} + \omega p_{31} + \omega^2 p_{12})$	p_{31}	p_{02}	$\omega^2 p_{12}$
$x'_1 =$	$-k(x_1 + x_2 + x_3)$	x_3	$-x_1$	$\omega^2 x_1$
$x'_2 =$	$-k(x_1 + \omega x_2 + \omega^2 x_3)$	x_1	$-x_3$	ωx_2
$x'_3 =$	$-k(x_1 + \omega^2 x_2 + \omega x_3)$	x_2	$-x_2$	ωx_3

$$* k = (\omega - \omega^2)/3 = 3^{-1/2}i, \quad \omega = e^{2\pi i/3}.$$

⁵⁸ Coxeter, 10, p. 181.

⁵⁹ Coxeter, 10, p. 178; 12, p. 477.

TABLE II

The group $[3^2, 2, 1]$, of order 51840, generated by anticollineations

	N_1	N	O	P	P_1	Q
$x'_1 =$	\bar{x}_1	$\omega \bar{x}_1$	$x_1 - s^*$	x_1	x_1	x_1
$x'_2 =$	x_2	x_2	$x_2 - s$	$\omega \bar{x}_2$	\bar{x}_2	x_2
$x'_3 =$	x_3	x_3	$x_3 - s$	x_3	x_3	$\omega \bar{x}_3$
$p'_{01} =$	$-p_{23}$	$-\omega p_{23}$	$p_{01} - t^\dagger$	p_{01}	p_{01}	p_{01}
$p'_{02} =$	p_{02}	p_{02}	$p_{02} - t$	$-\omega p_{31}$	$-p_{31}$	p_{02}
$p'_{03} =$	p_{03}	p_{03}	$p_{03} - t$	p_{03}	p_{03}	$-\omega p_{12}$
$p'_{23} =$	$-p_{01}$	$-\omega^2 p_{01}$	$p_{23} + t$	p_{23}	p_{23}	p_{23}
$p'_{31} =$	p_{31}	p_{31}	$p_{31} + t$	$-\omega^2 p_{02}$	$-p_{02}$	p_{31}
$p'_{12} =$	p_{12}	p_{12}	$p_{12} + t$	p_{12}	p_{12}	$-\omega^2 p_{03}$
$\bar{p}'_{01} =$	\bar{p}_{01}	$\omega \bar{p}_{01}$	$-\bar{p}_{23} - \bar{t}$	$-\bar{p}_{23}$	$-\bar{p}_{23}$	$-\bar{p}_{23}$
$\bar{p}'_{02} =$	$-\bar{p}_{31}$	$-\bar{p}_{31}$	$-\bar{p}_{31} - \bar{t}$	$\omega \bar{p}_{02}$	\bar{p}_{02}	$-\bar{p}_{31}$
$\bar{p}'_{03} =$	$-\bar{p}_{12}$	$-\bar{p}_{12}$	$-\bar{p}_{12} - \bar{t}$	$-\bar{p}_{12}$	$-\bar{p}_{12}$	$\omega \bar{p}_{03}$
$\bar{p}'_{23} =$	\bar{p}_{23}	$\omega^2 \bar{p}_{23}$	$-\bar{p}_{01} + \bar{t}$	$-\bar{p}_{01}$	$-\bar{p}_{01}$	$-\bar{p}_{01}$
$\bar{p}'_{31} =$	$-\bar{p}_{02}$	$-\bar{p}_{02}$	$-\bar{p}_{02} + \bar{t}$	$\omega^2 \bar{p}_{31}$	\bar{p}_{31}	$-\bar{p}_{02}$
$\bar{p}'_{12} =$	$-\bar{p}_{03}$	$-\bar{p}_{03}$	$-\bar{p}_{03} + \bar{t}$	$-\bar{p}_{03}$	$-\bar{p}_{03}$	$\omega^2 \bar{p}_{12}$
$z'_0 =$	\bar{z}_1	$\omega^2 \bar{z}_1$	$k(\bar{z}_1 + \bar{z}_2 + \bar{z}_3)$	$\omega^2 \bar{z}_2$	\bar{z}_2	$\omega^2 \bar{z}_2$
$z'_1 =$	$-\bar{z}_0$	$-\omega^2 \bar{z}_0$	$k(-\bar{z}_0 - \bar{z}_2 + \bar{z}_3)$	$-\omega \bar{z}_3$	$-\bar{z}_3$	$\omega \bar{z}_3$
$z'_2 =$	\bar{z}_3	$\omega \bar{z}_3$	$k(-\bar{z}_0 + \bar{z}_1 - \bar{z}_3)$	$-\omega^2 \bar{z}_0$	$-\bar{z}_0$	$-\omega \bar{z}_1$
$z'_3 =$	$-\bar{z}_2$	$-\omega \bar{z}_2$	$k(-\bar{z}_0 - \bar{z}_1 + \bar{z}_2)$	$\omega \bar{z}_1$	\bar{z}_1	$-\omega^2 \bar{z}_2$

$$* s = (x_1 + x_2 + x_3 + \bar{x}_1 + \bar{x}_2 + \bar{x}_3)/3.$$

$$\dagger t = (p_{01} + p_{02} + p_{03} - p_{23} - p_{31} - p_{12})/3.$$

TABLE III

New collineations for generating $[3^2, 2, 1]'$

	$U = N_1 N$	$V = N_1 O$	$W = N_1 P$	$X = N_1 P_1$	$Y = N_1 Q$
$x'_1 =$	$\omega^2 x_1$	$\bar{x}_1 - s^*$	\bar{x}_1	\bar{x}_1	\bar{x}_1
$x'_2 =$	x_2	$x_2 - s$	$\omega \bar{x}_2$	\bar{x}_2	x_2
$x'_3 =$	x_3	$x_3 - s$	x_3	x_3	$\omega \bar{x}_3$
$p'_{01} =$	$\omega^2 p_{01}$	$-p_{23} - t^\dagger$	$-p_{23}$	$-p_{23}$	$-p_{23}$
$p'_{02} =$	p_{02}	$p_{02} - t$	$-\omega p_{31}$	$-p_{31}$	p_{02}
$p'_{03} =$	p_{03}	$p_{03} - t$	p_{03}	p_{03}	$-\omega p_{12}$
$p'_{23} =$	ωp_{23}	$-p_{01} + t$	$-p_{01}$	$-p_{01}$	$-p_{01}$
$p'_{31} =$	p_{31}	$p_{31} + t$	$-\omega^2 p_{02}$	$-p_{02}$	p_{31}
$p'_{12} =$	p_{12}	$p_{12} + t$	p_{12}	p_{12}	$-\omega^2 p_{03}$
$z'_0 =$	$-\omega z_0$	$k(z_0 + z_2 - z_3)$	$-\omega^2 z_2$	$-z_3$	$\omega^2 z_2$
$z'_1 =$	$-\omega z_1$	$k(z_1 + z_2 + z_3)$	$-\omega z_2$	$-z_2$	$-\omega z_2$
$z'_2 =$	$-\omega^2 z_2$	$k(z_0 + z_1 - z_2)$	$\omega^2 z_1$	z_1	$-\omega z_0$
$z'_3 =$	$-\omega^2 z_3$	$k(-z_0 + z_1 - z_3)$	ωz_0	z_0	$\omega^2 z_1$

$$* s = (x_1 + x_2 + x_3 + \bar{x}_1 + \bar{x}_2 + \bar{x}_3)/3.$$

$$\dagger t = (p_{01} + p_{02} + p_{03} - p_{23} - p_{31} - p_{12})/3.$$

TABLE IV

Two ways of generating the group $[3^4, 2, 1]$, of order 696729600

	N_3	N_2	N_1	N	O	P	P_1	Q
$z'_0 =$	\tilde{z}_0	$z_0 - p^*$	z_0	z_0	z_0	z_0	z_0	z_0
$z'_1 =$	z_1	$z_1 + p$	\tilde{z}_1	$\omega \tilde{z}_1$	$z_1 - \sigma^\dagger$	z_1	z_1	z_1
$z'_2 =$	z_2	z_2	z_2	z_2	$z_2 - \sigma$	$\omega \tilde{z}_2$	\tilde{z}_2	z_2
$z'_3 =$	z_3	$z_3 - p$	z_3	z_3	$z_3 - \sigma$	z_3	z_3	$\omega \tilde{z}_3$
<hr/>								
$y'_0 = y_1$	$y'_1 = y_2$	$y'_2 = y_3$	$y'_3 = y_4$	$y'_4 = -y_5$	$y'_5 = y_6$	$y'_6 = y_7$	$y'_7 = y_\nu - \Sigma y/4$	
$y'_1 = y_0$	$y'_2 = y_1$	$y'_3 = y_2$	$y'_4 = y_3$	$y'_5 = -y_4$	$y'_6 = y_5$	$y'_7 = y_6$	$(\nu = 0, 1, \dots, 7)$	

$$* p = (z_0 - z_1 + z_3 + \omega \tilde{z}_0 - \omega \tilde{z}_1 + \omega \tilde{z}_3)/3.$$

$$\dagger \sigma = (z_1 + z_2 + z_3 - \tilde{z}_1 - \tilde{z}_2 - \tilde{z}_3)/3.$$

UNIVERSITY OF TORONTO.

REFERENCES.

1. H. F. Baker, *Principles of Geometry*, vol. 6, Cambridge, 1933.
2. W. W. Rouse Ball, *Mathematical Recreations and Essays*, eleventh edition, London, 1939.
3. H. F. Blichfeldt, *Finite Collineation Groups*, Chicago, 1917.
4. H. R. Brahana, "Pairs of generators of the known simple groups whose orders are less than one million," *Annals of Mathematics*, vol. 31 (1930), pp. 529-549.
5. R. Brauer and C. Nesbitt, "On the modular representations of groups of finite order," *University of Toronto Studies* (Mathematical Series, No. 4, 1937).
6. H. Burkhardt, "Untersuchungen aus dem Gebiete der hyperelliptischen Modul-functionen, III," *Mathematische Annalen*, vol. 41 (1893), pp. 313-343.
7. W. Burnside, "The determination of all groups of rational linear substitutions of finite order which contain the symmetric group in the variables," *Proceedings of the London Mathematical Society* (2), vol. 10 (1912), pp. 284-308.
8. H. S. M. Coxeter, "The pure Archimedean polytopes in six and seven dimensions," *Proceedings of the Cambridge Philosophical Society*, vol. 24 (1928), pp. 1-9.
9. —, "The polytopes with regular-prismatic vertex figures," Part I, *Philosophical Transactions of the Royal Society of London* (A), vol. 229 (1930), pp. 329-425.
10. —, Part 2, *Proceedings of the London Mathematical Society* (2), vol. 34 (1932), pp. 126-189.
11. —, "Discrete groups generated by reflections," *Annals of Mathematics*, vol. 35 (1934), pp. 588-621.
12. —, "Finite groups generated by reflections, and their subgroups generated by reflections," *Proceedings of the Cambridge Philosophical Society*, vol. 30 (1935), pp. 466-482.
13. —, "Wythoff's construction for uniform polytopes," *Proceedings of the London Mathematical Society* (2), vol. 38 (1935), pp. 327-339.

14. L. E. Dickson, *Linear Groups, with an exposition of the Galois Field Theory*, Leipzig, 1901.*
15. P. Du Val, "On the directrices of a set of points in a plane," *Proceedings of the London Mathematical Society* (2), vol. 35 (1932), pp. 23-74.
16. E. L. Elte, *The Semi-regular Polytopes of the Hyperspaces*, Groningen, 1912.
17. J. S. Frame, "The simple group of order 25920," *Duke Mathematical Journal*, vol. 2 (1936), pp. 477-484.
18. ———, "A symmetric representation of the twenty-seven lines on a cubic surface by lines in a finite geometry," *Bulletin of the American Mathematical Society*, vol. 44 (1938), pp. 658-661.
19. T. Gosset, "On the regular and semi-regular figures in space of n dimensions," *Messenger of Mathematics*, vol. 29 (1900), pp. 43-48.
20. A. Henderson, *The Twenty-seven Lines upon the Cubic Surface*, Cambridge, 1911.
21. H. Maschke, "Aufstellung des vollen Formensystems einer quaternären Gruppe von 51840 linearen Substitutionen," *Mathematische Annalen*, vol. 33 (1888), pp. 317-344.
22. L. Schläfli, "An attempt to determine the twenty-seven lines upon a surface of the third order, and to divide such surfaces into species in reference to the reality of the lines upon the surface," *Quarterly Journal of Mathematics*, vol. 2 (1858), pp. 110-120.
23. P. H. Schoute, *Mehrdimensionale Geometrie*, vol. 1, Leipzig, 1902.
24. ———, "On the relation between the vertices of a definite six-dimensional polytope and the lines of a cubic surface," *Koninklijke Akademie van Wetenschappen te Amsterdam, Proceedings of the Section of Sciences*, vol. 13 (1910), pp. 375-383.
25. H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Leipzig, 1934.
26. D. M. Y. Sommerville, *An Introduction to the Geometry of n dimensions*, London, 1929.
27. J. Steiner, "Über die Flächen dritten Grades," *Journal für die reine und angewandte Mathematik*, vol. 53 (1857), pp. 133-141.
28. J. A. Todd, "Polytopes associated with the general cubic surface," *Journal of the London Mathematical Society*, vol. 7 (1932), pp. 200-205.
29. A. Witting, *Ueber eine der Hesse'schen Configuration der ebenen Curve dritter Ordnung analoge Configuration im Raume, auf welche die Transformationstheorie der hyperelliptischen Functionen ($p = 2$) führt*, Dresden, 1887.

LE MOUVEMENT BROWNIEN PLAN.*

Par M. PAUL LÉVY.

Introduction. Nous nous proposons d'étudier les fonctions aléatoires $X(t)$, $Y(t)$ qui, dans le mouvement brownien, et en projection sur un plan, définissent les coordonnées du centre d'une molécule considérée indépendamment des autres. Il s'agira seulement du *mouvement brownien mathématique*, dont nous donnerons plus loin la définition précise. Indiquons tout de suite qu'il se distingue du mouvement brownien réel parce que le libre parcours moyen est supposé réduit à zéro. On est ainsi conduit à étudier des fonctions continues d'allure excessivement irrégulières: elles présentent, dans tout intervalle, une infinité de maxima et minima; leur quatre nombre dérivés sont infinis, sauf sur un ensemble de mesure nulle où un de ces nombres est fini et sur un ensemble dénombrable où deux sont finis; il y en a en tout point au moins deux qui sont infinis. Il s'agit là bien entendu de propriétés presque sûres, c'est-à-dire réalisées avec une probabilité unité, mais pouvant être en défaut dans des cas possibles, quoique de probabilité nulle; il nous arrivera de ne pas rappeler la nécessité de cette restriction; il nous semble que, dans des questions où il est bien entendu qu'il s'agit de phénomènes aléatoires, il ne peut en résulter aucune ambiguïté.

L'indépendance stochastique de $X(t)$ et $Y(t)$ a conduit les mathématiciens à étudier d'abord le mouvement brownien linéaire (c'est-à-dire projeté sur une droite). Cette étude a pris naissance dans les travaux indépendants les uns des autres de Bachelier¹ et de N. Wiener,² et de nombreux travaux lui ont été consacrés depuis quelques années. Le lecteur pourra trouver un exposé des principes fondamentaux de la théorie ainsi édifiée dans

* Received October 17, 1939.

¹ Bachelier, *Calcul des probabilités* (1912). A cette date, Bachelier apparaît comme un précurseur. Si la manière dont sont introduits les problèmes où le temps joue le rôle d'une variable continue laisse à désirer, il n'en reste pas moins que c'est dans cet ouvrage que l'on trouve pour la première fois l'idée que la loi de Gauss s'introduit nécessairement comme conséquence de la continuité d'un processus additif, et la relation entre ce processus et l'équation de la chaleur. Il faut aussi signaler plusieurs formules relatives à l'écart maximum, et peut-être la formule (que j'ai cherchée en vain dans un grand nombre d'ouvrages antérieurs) qui, dans le cas des lois absolument continues, définit la loi dont dépend la somme de deux variables aléatoires indépendantes.

² N. Wiener, *Differential Space* (Publications of the Massachusetts Institute of Technology, Ser. II, N° 60, juin 1923).

notre *Théorie de l'addition des variables aléatoires* (1937), p. 166 à 173. Des résultats nouveaux ont été indiqués dans notre travail récent *sur quelques processus stochastiques homogènes*.³ Nous rappellerons au § 1 du présent travail la définition mathématique du mouvement brownien, c'est-à-dire la définition stochastique de $X(t)$, et quelques propriétés connues de ce mouvement utiles pour la suite. Quant aux théorèmes généraux du calcul des probabilités que nous considérerons comme connus, ils se trouvent tous dans notre ouvrage de 1937 cité ci-dessus; nous prions le lecteur de s'y reporter en cas de besoin.

Les § 2 et 3, concernant respectivement le mouvement brownien linéaire et le mouvement plan, sont consacrés à l'étude de quelques propriétés locales des trajectoires. Les § 4 et 5 sont consacrés respectivement à l'étude d'une expression qu'on peut représenter symboliquement par l'intégrale

$$B = \int_0^1 [dX(t)]^2$$

et que nous appelons *l'oscillation brownienne* de $X(t)$ dans l'intervalle $(0, 1)$, et à celle de l'intégrale

$$S = \int_0^1 Y(t) dX(t),$$

qui, avec des axes convenablement choisis, représente l'aire comprise entre la courbe C trajectoire du mouvement brownien plan pendant l'intervalle de temps $(0, 1)$, et sa corde.

Il s'agit là d'intégrales stochastiques d'un type essentiellement nouveau. On peut les définir comme limites de sommes, ce qui, pour l'aire S , revient à la considérer comme limite de celle définie en remplaçant la courbe C par une ligne polygonale inscrite; mais il ne s'agit pas d'une limite ordinaire; suivant les cas il s'agira de convergence en probabilité, ou de convergence en moyenne quadratique (qui, comme on sait, implique la précédente), ou de convergence presque sûre. Cette dernière notion ne peut d'ailleurs intervenir que comme conséquence d'une hypothèse restrictive relative au mode de division de l'intervalle de variation de t en intervalles très nombreux et très petits; il nous suffira de supposer que tout point de division une fois choisi soit conservé dans les subdivisions ultérieures.

Le hasard peut d'ailleurs intervenir, d'une part dans le choix des points de division, d'autre part dans celui des fonctions $X(t)$ et $Y(t)$. Le point

³ *Compositio Mathematica*, vol. 7 (1939), pp. 283-339. Ce travail sera désigné dans la suite par l'abréviation "processus," et notre livre cité dans le texte sera désigné par "Var. aléatoires."

de vue le plus simple consiste à fixer un mode de division de l'intervalle d'intégration, et à considérer $X(t)$ et $Y(t)$ comme aléatoires, et démontrer dans ces conditions l'existence des limites B et S ; c'est ce que nous ferons au début de chacun des § 4 et 5. Mais nous envisagerons ensuite le point de vue inverse; c'est de ce point de vue que nous considérons qu'il introduit en analyse une notion d'une nature toute nouvelle. Pour chaque détermination, soit de $X(t)$, soit de la courbe C , les points de division étant choisis au hasard, il peut arriver que les sommes utilisées pour définir B et S convergent presque sûrement vers des limites, et cela sans qu'on puisse conclure à l'existence de l'intégrale au sens ordinaire, c'est-à-dire d'une limite indépendante du mode de division de l'intervalle d'intégration et existant dans tous les cas. Il s'agit là de propriétés non aléatoires de chaque fonction $X(t)$ ou de chaque courbe C . Le résultat peut-être le plus important de ce travail est que, dans le schéma aléatoire du mouvement brownien mathématique, on obtient presque sûrement des trajectoires ayant ces propriétés. Quant à la nature de la limite, dans le cas de B , elle n'est pas aléatoire; on a $B = 1$. Dans le cas de S , c'est une nouvelle variable aléatoire dont nous avons cherché à définir la nature; c'est l'objet de la fin du § 5 (5° à 12° de ce paragraphe). Il ne semble pas que la fonction de répartition de S soit susceptible d'une expression simple; nous donnons plusieurs résultats relatifs à cette fonction, dont le plus important est peut-être le suivant: la détermination de la loi à deux variables S et L (L étant la longueur de la corde sous-tendant l'arc C) dépend d'une équation aux dérivées partielles du second ordre et du type elliptique, vérifiée par une des dérivées de la fonction de répartition, et qui, compte tenu de ce qu'il s'agit d'une fonction de répartition, la détermine complètement.

Observons, en ce qui concerne B , que le fait que le hasard réalise avec une probabilité unité des fonctions pour lesquelles cette intégrale existe, n'implique pas qu'il soit facile de nommer une telle fonction. Le § 4 contient, sur ce sujet, quelques remarques qu'il peut être utile de compléter et préciser. On pourrait étudier aussi l'aire S au même point de vue, en cherchant à nommer une courbe pour laquelle l'aire existe au point de vue stochastique, mais non au point de vue de l'analyse ordinaire.

Au § 6, nous montrerons que la courbe C est, avec une probabilité unité, un ensemble de mesure superficielle nulle. Pourtant le fait que B soit positif implique que la courbe fasse assez de détours infiniment petits pour pouvoir remplir une aire. Mais, pour qu'elle remplisse effectivement une aire, il faudrait une organisation de ces détours infiniment petits que le hasard n'a aucune chance de produire.

Cette remarque s'étend à d'autres schémas aléatoires que celui du mouvement brownien. Au § 7, nous étudions quelques exemples de tels schémas. Tous vérifient cette condition que la courbe Γ décrite par le point mobile quand le paramètre t varie de zéro à un est composée de deux arcs stochastiquement semblables à la courbe complète; nous entendons par là que n'importe quel ensemble de courbes qui sont des courbes Γ possibles, est aussi, à une similitude près, un ensemble de formes possibles de chacun de ces arcs, et cela avec la même probabilité que pour Γ . Cette propriété, dans le cas du mouvement brownien, est réalisée pour l'arc correspondant à n'importe quel intervalle de variation de t , et le rapport de similitude stochastique est la racine carrée de la longueur de cet intervalle. Dans le cas des schémas étudiés au § 7, cette propriété n'appartient qu'aux arcs correspondant aux intervalles de variation de t compris entre deux multiples consécutifs de 2^{-n} , de sorte que l'ensemble des valeurs de t qui sont des fractions dyadiques joue, dans l'étude de ces courbes, un rôle tout à fait remarquable. Le rapport de similitude stochastique, suivant les schémas étudiés, est déterminé pour chacun des arcs partiels considérés, ou au contraire aléatoire; mais dans ce cas les rapports relatifs aux différents arcs ne sont pas indépendants. Toutes les fois que la somme des carrés de ces rapports, étendue à n'importe quelle division de la courbe en arcs stochastiquement semblables à la courbe entière, est égale à l'unité, on peut dire, comme dans le cas du mouvement brownien, que la courbe fait assez de détours infiniment petits pour pouvoir remplir une aire. Il peut alors s'agir de courbes non aléatoires, composées de parties semblables au tout, et notamment de deux courbes bien connues qui seront désignées par Γ_0 et Γ_1 , et qui remplissent effectivement des aires. Mais, toutes les fois que le hasard joue un rôle suffisant dans la définition de la courbe, pour les mêmes raisons que dans le cas du mouvement brownien, il est infiniment peu probable que la courbe remplisse effectivement une aire.

Nous étudions aussi, pour ces schémas, l'aire S comprise entre l'arc et la courbe. Dans les cas où nous venons d'indiquer que la courbe pourrait à première vue remplir une aire, mais n'a aucune chance de le faire effectivement, cette aire se présente sous la forme d'une somme $\sum \pm s_v$ d'aires triangulaires, la série $\sum s_v^2$ étant convergente. Si alors les signes sont choisis au hasard, la série qui définit S est presque sûrement convergente; il est presque sûr que la courbe étudiée limite une aire définie au sens indiqué à propos du mouvement brownien. Si au contraire tous les termes sont positifs, ou si (ce qui sera le cas pour les schémas qui seront désignés par les notations Γ_2 et Γ'_2) ils sont groupés en groupes étendus de termes de même signe, la série considérée est divergente; S apparaît alors comme infini ou in-

déterminé. Comme le signe des aires triangulaires $\pm s_v$ peut être défini par une loi précise même dans des cas où le hasard intervient par ailleurs dans la définition de la courbe, la question de l'existence de l'aire S n'est pas liée à la précédente; du moins l'existence de l'aire S ne permet aucune conclusion au sujet de la mesure superficielle de l'ensemble des points de la courbe.

Le grand nombre des schémas aléatoires que l'on pourrait ainsi étudier et des problèmes qui se posent nous a obligé, non seulement à faire un choix, mais dans certains cas à nous contenter de démonstrations résumées ou même d'énoncés sans démonstrations, et aussi à énoncer sans en indiquer la solution des problèmes qui nous semblent mériter d'être étudiés ultérieurement.

Le plan initial de ce travail comportait un dernier paragraphe consacré à des types d'intégrales et d'équations différentielles stochastiques qui généralisent l'aire S , et notamment à l'intégrale

$$U = \int \phi(X, Y, Z) dX(t) + \chi(X, Y, Z) dY(t) + \psi(X, Y, Z) dZ(t)$$

qui définit le travail de la particule mobile dans un champ de forces, et à l'équation aux différentielles totales

$$dU(t) = \phi(X, Y, Z, U) dX(t) + \chi(X, Y, Z, U) dY(t) + \psi(X, Y, Z, U) dZ(t),$$

X, Y, Z étant les coordonnées d'une particule dans le mouvement brownien à trois dimensions, et ϕ, χ, ψ étant des fonctions continues à la Lipschitz. On peut définir U comme limite de sommes ou de solutions d'équations aux différences finies. Le point de vue auquel nous envisageons l'étude de ces équations est d'ailleurs différent de celui adopté par M. S. Bernstein dans son mémoire connu sur les équations différentielles stochastiques en ce sens que nous supposons les expériences qui déterminent X, Y, Z effectuées avant l'intégration. L'intégrale $U(t)$ ayant une valeur initiale donnée sera donc une fonctionnelle dépendant d'une manière non aléatoire des trois fonctions aléatoires X, Y, Z . En vertu de propriétés presque sûres de ces fonctions, les opérations qui aboutissent à la définition de $U(t)$ ont un sens, dans les conditions indiquées à propos des expressions B et S : il peut exister des modes de division de l'intervalle d'intégration en intervalles partiels pour lesquelles il n'y ait pas convergence de ces opérations; mais ces modes de division sont exceptionnels.

On peut naturellement généraliser la théorie précédente en prenant pour la trajectoire du point X, Y, Z un schéma aléatoire différent de celui du mouvement brownien.

Les circonstances présentes (en septembre 1939) m'ont décidé à renoncer,

pour le moment, à la rédaction de cette dernière partie, et à demander à la direction de l'*American Journal of Mathematics* de publier ce travail dans son état actuel; je l'en remercie à l'avance.⁴

1. Définition de la fonction aléatoire $X(t)$. Nous considérerons une variable réelle t , variant de zéro à l'infini, et désignerons par $X(t)$ une fonction aléatoire de t ayant les caractères suivants: $X(0) = 0$; *quels que soient $t \geq 0$ et $\tau > 0$, l'accroissement $\Delta X(t) = X(t + \tau) - X(t)$ est une variable gaussienne d'écart type $\sqrt{\tau}$; il est de plus stochastiquement indépendant du passé* [c'est-à-dire de l'ensemble des valeurs prises par $X(t')$ dans l'intervalle $(0, t)$].

Rappelons comment on peut définir une suite d'expériences aboutissant à la détermination d'une fonction aléatoire $X(t)$ ayant effectivement les caractères précédents. Considérons à cet effet trois valeurs t_0, t_1, t_2 de t ($t_0 < t_1 < t_2$), et posons

$$\begin{aligned} X' &= X(t_1) - X(t_0), & X'' &= X(t_2) - X(t_1), \\ X &= X' + X'' = X(t_2) - X(t_0). \end{aligned}$$

Les conditions imposées à $X(t)$ ne sont évidemment compatibles que parce que la somme des deux variables gaussiennes X' et X'' , d'écart types respectifs $\sqrt{t_1 - t_0}$ et $\sqrt{t_2 - t_1}$, est bien une variable gaussienne d'écart type $\sqrt{t_2 - t_0}$. Dans ces conditions, quand on connaît $X(t_0)$, on dispose de deux procédés stochastiquement équivalents pour déterminer $X(t_1)$ et $X(t_2)$. On peut, par deux expériences indépendantes, déterminer les deux accroissements successifs X' et X'' . On peut aussi déterminer d'abord l'accroissement total, puis *interpoler*, c'est-à-dire déterminer $X(t_1)$ d'après la loi de probabilité conditionnelle dont dépend cette variable lorsque $X(t_0)$ et $X(t_2)$ sont connus. Dans ces conditions, $X(t_1)$ a pour valeur probable le nombre

$$m_1 = \frac{(t_2 - t_1)X(t_0) + (t_1 - t_0)X(t_2)}{t_2 - t_0}$$

obtenu par une interpolation linéaire, et $X(t_1) - m_1$ est une variable gaussienne d'écart type

$$\sigma_1 = \sqrt{\frac{(t_1 - t_0)(t_2 - t_1)}{t_2 - t_0}}.$$

En particulier, si $t_1 - t_0 = t_2 - t_1 = \tau$, on a $\sigma_1 = \sqrt{\tau/2}$, c'est-à-dire que la différence entre $X(t_1)$ et sa valeur probable est une variable gaussienne d'écart type $\sqrt{2}$ fois plus petit que quand on connaît seulement $X(t_0)$.

⁴ Quelques-uns des résultats établis dans ce travail ont été énoncés dans deux Notes présentées à l'Académie des Sciences (*C. R.*, t. 207, p. 1152; t. 209, p. 140, et erratum, p. 387).

Ce résultat est en relation évidente avec les propriétés connues de la loi de Gauss à deux variables, dans le cas isotrope: d'après ces propriétés, si X' et X'' sont deux variables gaussiennes indépendantes de même écart type $\sqrt{\tau}$,

$$\frac{X}{2} = \frac{X' + X''}{2} \quad \text{et} \quad \frac{X' - X''}{2} = X' - \frac{X}{2}$$

sont deux variables gaussiennes indépendantes d'écart type $\sqrt{\tau/2}$; cette remarque définit parfaitement la loi dont dépend X' lorsque X est connu.

On peut donc, sans changer la loi de probabilité imposée pour l'ensemble des trois variables $X(t_0)$, $X(t_1)$, et $X(t_2)$, procéder par interpolation, c'est-à-dire déterminer $X(t_1)$ seulement après la détermination préalable de $X(t_0)$ et $X(t_2)$. Par suite, pour déterminer $X(t)$ dans l'intervalle $(0, 1)$, on peut déterminer $X(1)$, puis, par des interpolations successives, déterminer $X(\frac{1}{2})$, puis $X(\frac{1}{4})$ et $X(\frac{3}{4})$, et ainsi de suite. Désignons par $X_n(t)$ la fonction continue égale aux valeurs ainsi obtenues quand t est multiple de 2^{-n} , et variant linéairement dans chacun des intervalles compris entre deux multiples consécutifs de ce nombre; $X_n(t)$ peut être considéré comme la $n^{\text{ième}}$ approximation de $X(t)$, et $X(t)$ peut être défini comme la limite presque sûre de ces approximations. On a en effet le théorème suivant:

THÉORÈME 1. *Il y a une probabilité unité pour que, quand n augmente indéfiniment, $X_n(t)$ tende vers une fonction continue $X(t)$, et cela uniformément dans l'intervalle $(0, 1)$.*

La démonstration est très simple. Désignons par δ_n le maximum de $|X_{n+1}(t) - X_n(t)|$, dans l'intervalle $(0, 1)$. C'est le plus grand des modules de 2^n variables gaussiennes de même écart type q^{n+2} ($q = 1/\sqrt{2}$). Compte tenu du lemme de Boole, on a donc

$$(1) \quad \alpha_n = P_r\{\delta_n \geq q^{n+2}x_n\} \leq \frac{2^{n+1}}{x_n \sqrt{2\pi}} e^{-x_n^2/2}.$$

Si l'on prend pour x_n la valeur $c\sqrt{2n \log 2}$, avec $c > 1$, α_n est le terme général d'une série convergente. D'après le lemme de M. Cantelli (ou celui de M. Borel, puis-que les δ_n sont indépendants), il existe alors presque sûrement un nombre fini N tel que, pour $n > N$, on ait $\delta_n < cq^{n+1}\sqrt{n \log 2}$, ce qui établit la convergence uniforme presque sûre de $X_n(t)$ vers une limite, évidemment continue, $X(t)$, c. q. f. d.

Bien entendu le nombre N est aléatoire. Si donc la convergence est uniforme par rapport à $X(t)$ (de zéro à un, ou dans n'importe quel intervalle fini), elle n'est pas uniforme par rapport au choix de $X(t)$; mais il suffit d'écarter des cas de probabilité arbitrairement petite pour qu'elle le devienne.

On vérifie aisément que la fonction $X(t)$ ainsi obtenue vérifie toutes les conditions indiquées dans la définition théorique. On peut d'autre part arriver au même résultat en remplaçant l'ensemble des fractions dyadiques, qui joue un rôle essentiel dans les expériences que nous venons de définir, par n'importe quel ensemble de valeurs de t dénombrable et partout dense dans l'intervalle où l'on veut définir $X(t)$. Le résultat obtenu est stochastiquement indépendant du choix de cet ensemble.

Rappelons maintenant trois lemmes dont le lecteur trouvera la démonstration dans nos travaux antérieurs [*Processus*, formules (15) et (20); *Var. aléatoires*, p. 172]. Le premier remonte à Bachelier.

LEMME 1. Pour $x > 0$, on a

$$Pr\left\{\max_{0 \leq u \leq t} X(u) > x\right\} = Pr\{|X(t)| > x\} = \sqrt{2/\pi t} \int_x^\infty e^{-u^2/2t} du.$$

LEMME 2. Pour x supérieur à la fois à 0 et a , on a

$$Pr\left\{\max_{0 \leq u \leq t} X(u) > x/X(t) = a\right\} = e^{-2x(x-a)/t}.$$

[Rappelons que $Pr\{A/B\}$ désigne la probabilité de A dans l'hypothèse B .]

LEMME 3. Etant donnés $t_1 > 0$ et $c > 1$, il existe presque sûrement un nombre $\eta > 0$ tel que, pour $t \leq t_1$ et $0 < \tau \leq \eta$, on ait

$$|X(t + \tau) - X(t)| < c\sqrt{2\tau \log 1/\tau}.$$

Si au contraire $c < 1$, la probabilité de l'existence de η est nulle.

2. Etude locale de $X(t)$. Nous pouvons nous contenter d'étudier les propriétés de $X(t)$ au voisinage de l'origine; les résultats obtenus s'appliqueront évidemment au voisinage de n'importe quel autre point, soit à droite, soit à gauche de ce point; mais bien entendu, si l'on trouve qu'une certaine circonstance est infiniment peu probable au voisinage de n'importe quel point donné d'avance, cela n'empêche pas qu'il puisse exister avec une probabilité positive des points, impossibles à connaître à l'avance, au voisinage desquels elle soit réalisée.

Commençons par établir un théorème qui ramène l'étude locale de $X(t)$ à l'étude asymptotique de cette fonction, pour t infini. Pour l'énoncer, nous poserons

$$(2) \quad t = e^u, \quad X(t) = \sqrt{t} \phi(u),$$

de sorte que $\phi(u)$ est, pour chaque valeur de u , une variable gaussienne réduite.

THÉORÈME 2. *La définition stochastique de $\phi(u)$ est invariante par le changement de u en $-u$ (évidemment aussi par celui de u en $u+c$).*

En d'autres termes, les propriétés stochastiques de $X(t)/\sqrt{t}$ sont invariantes par le changement de t en $1/t$.

Considérons d'abord la suite des valeurs $t_n = q^n$ de t , q étant ici un nombre quelconque entre zéro et un; posons $X(t_n) = X_n = \phi_n \sqrt{t_n}$. De l'indépendance des accroissements successifs de $X(t)$, on déduit

$$(3) \quad \phi_{n-1} = \phi_n \sqrt{q} + \phi'_n \sqrt{1-q},$$

ϕ'_n étant une variable gaussienne réduite indépendante de $\phi_n, \phi_{n+1}, \dots$; d'autre part, d'après le principe d'interpolation exposé au § 1, $X(t_{n+1})$ pouvant être déterminé par interpolation entre $X(0)$ et $X(t_n)$, on a

$$(4) \quad \phi_{n+1} = \phi_n \sqrt{q} + \phi''_n \sqrt{1-q},$$

ϕ''_n étant une variable gaussienne réduite indépendante de $\phi_n, \phi_{n-1}, \dots$; on a donc les mêmes expériences à faire pour déterminer la suite des ϕ_n de gauche à droite, ou bien de droite à gauche. En d'autres termes la nature stochastique de cette suite est invariante par le changement de n en $-n$ (ou en $h-n$, h étant un entier donné).

Cette symétrie stochastique n'est d'ailleurs pas détruite quand, après avoir déterminé $\phi(n)$ (c'est-à-dire ϕ_{-n} , si on a pris pour q la valeur $1/e$) pour toutes les valeurs entières de n , on effectue des interpolations pour déterminer les nombres $\phi(n + \frac{1}{2})$; chaque nombre $\phi(n + \frac{1}{2})$ a en effet la même corrélation avec $\phi(n)$ et avec $\phi(n+1)$. Le résultat obtenu n'est d'ailleurs autre que la suite des ϕ_n pour $q = e^{-\frac{1}{2}}$. En effectuant de nouvelles interpolations, on arrivera à définir $\phi(u)$ pour les valeurs de u multiples de $\frac{1}{4}$, puis de $\frac{1}{8}$, et ainsi de suite. A chacune de ces opérations, la symétrie est conservée, et l'on aboutit, à la limite, à la détermination de $\phi(u)$ par un processus stochastique absolument symétrique. Or il équivaut bien à celui par lequel nous avons défini $\phi(u)$, puisque, dans la définition de $X(t)$, rien n'empêche de choisir, pour les interpolations, les valeurs de t dont les logarithmes sont les valeurs de u qui interviennent dans le procédé de détermination de $\phi(u)$ que nous venons de décrire. Le théorème 2 est ainsi démontré.

Il entraîne bien simplement de nombreuses conséquences. Ainsi l'ensemble des racines de $X(t)$ n'a presque sûrement aucune borne supérieure; cela résulte aisément de ce que, d'après le lemme 1 appliqué à $x = X(u)$, on peut déterminer une fonction $f(t, x)$ supérieure à t et telle que, dans l'hypothèse $X(t) = x$, $X(u)$ ait, avec une probabilité donnée α , au moins une racine comprise entre t et $f(t, x)$. En prenant alors pour τ_0 un nombre

arbitraire, et posant $\tau_{n+1} = f[\tau_n, X(\tau_n)]$, on obtient une suite de nombres aléatoires croissants qui séparent des intervalles dont chacun a une probabilité α de contenir une racine de $X(t)$. Ces probabilités étant indépendantes, il résulte de la loi forte des grands nombres, sous sa forme la plus classique, qu'il y a presque sûrement une infinité d'intervalles (τ_n, τ_{n+1}) , ayant une fréquence tendant vers α , contenant chacun au moins une racine de $X(t)$.

Il résulte alors du théorème 2 que les racines positives de $X(t)$ ne sont pas non plus bornées inférieurement: *il est presque sûr que la racine zéro de $X(t)$ n'est pas isolée.*

Bien entendu, comme toutes les fois que l'on applique une méthode de transformation, on peut transformer, non seulement le théorème, mais sa démonstration, et obtenir ainsi une démonstration indépendante de la transformation employée. Dans le cas qui nous occupe, c'est ce qui donne à la fois la démonstration la plus simple du résultat obtenu, et celle qui permet le mieux son extension à d'autres types de fonctions aléatoires.

L'ensemble des racines de $X(t)$ étant fermé, et ne comprenant aucun intervalle, il existe bien entendu des racines isolées à gauche, ou à droite; elles forment au plus un ensemble dénombrable. Si nous définissons une racine θ par la condition d'être la plus petite racine au moins égale à un nombre donné t_0 , sauf peut-être dans l'hypothèse infiniment peu probable $X(t_0) = 0$, elle est isolée à gauche; ce renseignement ne modifiant en rien l'allure probable de $X(t)$ pour $t > t_0$, il est presque sûr que θ est, comme zéro, une racine non isolée à droite. Or il y a au plus une infinité dénombrable de racines isolées à gauche, donc d'occasions de trouver une racine isolée à la fois à gauche et à droite, et chaque fois la probabilité de cette circonstance est nulle. La probabilité totale est donc nulle: *il n'y a presque sûrement aucune racine de $X(t)$ isolée à la fois à gauche et à droite.*

Naturellement, ce résultat s'applique aux racines de $X(t) = x$, si x est donné. Pourtant $X(t)$ a dans tout intervalle une infinité dénombrable de maxima et minima; mais l'ensemble des valeurs maxima ou minima de $X(t)$ n'a aucune chance de contenir une valeur x donnée d'avance.

Pour une étude plus complète de l'ensemble des racines de $X(t)$, le lecteur peut se reporter à notre mémoire antérieur (*Processus*, § 7).

Indiquons une autre application du théorème 2. M. Khintchine a établi le théorème du logarithme itéré, d'après lequel

$$(5) \quad \Pr \left\{ \limsup_{|u| \rightarrow \infty} \frac{|\phi(u)|}{\sqrt{2 \log |u|}} = 1 \right\} = 1.$$

Il l'a démontré successivement pour u tendant vers $+\infty$, et pour u tendant vers $-\infty$ (donc t vers zéro). D'après le théorème 2, un de ces résultats entraîne l'autre.

Nous nous proposons maintenant, en supposant toujours, pour fixer les idées, $q < 1$, d'étudier quelques propriétés de la suite des nombres $X_n = \phi_n \sqrt{t_n}$. D'après la relation de récurrence (4), ils forment une chaîne de Markoff simple et *homogène dans le temps* (c'est-à-dire *stationnaire*; mais ce terme, généralement employé, nous paraît impropre). Nous pourrions renvoyer le lecteur à la théorie générale de ces chaînes; mais l'étude précise d'un cas particulier n'est peut-être pas inutile.

Indiquons d'abord la formule

$$(5') \quad \Pr \left\{ \limsup_{n \rightarrow \infty} \frac{|\phi(n)|}{\sqrt{2 \log n}} = 1 \right\} = 1,$$

analogue à la formule (5). Il nous suffira, pour la suite, de savoir que la limite considérée ne saurait dépasser l'unité, c'est-à-dire que, si $c > 1$, il existe presque sûrement un nombre N tel que, pour $n > N$, on ait

$$(6) \quad |\phi_n| < c\sqrt{2 \log n}.$$

Cela résulte immédiatement de ce que la probabilité de l'inégalité inverse, égale à

$$\sqrt{\frac{2}{\pi}} \int_{c\sqrt{2 \log n}}^{\infty} e^{-u^2/2} du < \sqrt{\frac{2}{\pi}} \int_{c\sqrt{2 \log n}}^{\infty} e^{-u^2/2} \frac{udu}{c\sqrt{2 \log n}} = \frac{1}{cn^c \sqrt{\pi \log n}},$$

est le terme général d'une série convergente.

Démontrons maintenant que la loi forte des grands nombres s'applique à la suite des $|\phi_n|$, c'est-à-dire que :

THÉORÈME 3. *La fréquence des ϕ_n inférieurs à un nombre donné x tend presque sûrement (et cela d'une manière uniforme) vers la probabilité théorique*

$$\Pr\{\phi_n < x\} = F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-a^2/2} da.$$

Nous montrerons que la différence entre cette fréquence et $F(x)$ devient presque sûrement inférieure en valeur absolue à un nombre donné arbitrairement petit que nous désignerons par 3ϵ . Il suffit évidemment d'établir le résultat analogue à celui énoncé, mais ne faisant intervenir que les valeurs des ϕ_n pour les valeurs de n de la forme $n_0 + \nu p$ ($\nu = 0, 1, 2, \dots$); p est un entier que rien n'empêche de choisir en fonction de ϵ et aussi grand qu'il est nécessaire, et n_0 est un quelconque des nombres $1, 2, \dots, p$. L'idée directrice de la démonstration est que, si p est assez grand, ϕ_n et ϕ_{n+p} sont presque indépendants, et qu'on peut appliquer la loi forte des grands nombres comme s'il s'agissait d'une suite de variables aléatoires indépendantes.

D'après la formule (4), que l'on peut appliquer à l'étude de la corrélation entre ϕ_n et ϕ_{n+h} à condition de remplacer q par q^h , on a

$$(7) \quad \phi_{n+h} = \phi_n \sqrt{q^h} + \psi_n \sqrt{1 - q^h},$$

ψ_n étant une variable gaussienne réduite, indépendante de ϕ_n . La corrélation entre ϕ_n et ϕ_{n+h} est bien, d'après cela, d'autant plus faible que h est plus grand. En désignant par ϕ un nombre positif donné, nous choisirons h assez grand pour que

$$(8) \quad \phi \sqrt{q^h} \leq \epsilon, \quad \sqrt{1 - q^h} \geq 1 - \epsilon.$$

Si alors $|\phi_n| = \phi$, il résulte de la formule (7) que

$$(9) \quad |\phi_{n+h} - \psi_n| \leq \epsilon(1 + |\psi_n|),$$

d'où l'on déduit aisément, du moins si ϵ est assez petit

$$(10) \quad |Pr\{\phi_{n+h} < x / |\phi_n| = \phi\} - F(x)| \leq \epsilon.^5$$

Si ψ désigne une variable gaussienne réduite, l'expression

$$(11) \quad \Omega = \frac{1}{|\log q|} \log \text{Max} \left[\left(1 + \frac{1 + \epsilon}{\epsilon} |\psi| \right)^2, \frac{1}{2\epsilon - \epsilon^2} \right],$$

où $\text{Max}[a, b]$ désigne le plus grand des nombres a et b , est une variable aléatoire à valeur probable finie. En désignant sa fonction de répartition par $G(\omega)$, on peut donc prendre pour p un entier tel que

$$\int_p^\infty \omega dG(\omega) \leq \epsilon,$$

et, en désignant par H le plus petit multiple de p au moins égal à Ω , sa valeur probable est

$$\mathcal{E}\{H\} = \int_0^\infty H dG(\omega) < \int_0^p p dG(\omega) + \int_p^\infty (p + \omega) dG(\omega) \leq p + \epsilon.$$

Désignons alors par $\psi'_0, \psi'_1, \dots, \psi'_k, \dots$ des déterminations indépendantes de ψ , et, par H_{k+1} la détermination de H qui correspond à ψ'_k .

⁵ Il n'y a qu'à utiliser cette remarque évidente que, si $\left| \log \frac{\psi'}{\psi} \right| < \epsilon_1$ (avec $\psi\psi' > 0$), et $|\psi'' - \psi'| < \epsilon_2$, et si, $f(x)$ étant la densité de probabilité de ψ , on a $\alpha |x| f(x) < m_1$ et $f(x) < m_2$, on commet sur la fonction de répartition une erreur au plus égale à $\epsilon_1 m_1 + \epsilon_2 m_2$ en remplaçant ψ par ψ'' . Dans le cas de la loi de Gauss, $m_1 = m_2 = \frac{1}{\sqrt{2\pi}}$; si ϵ est assez petit, on peut prendre $\epsilon_2 = \epsilon$ et $\epsilon_1 = (\sqrt{2\pi} - 1)\epsilon$; la formule (10) en résulte.

D'après la loi forte des grands nombres (qui s'applique à H , d'après un théorème de A. Kolmogoroff), on a presque sûrement

$$(12) \quad \lim_{k \rightarrow \infty} \frac{H_1 + H_2 + \dots + H_k}{k} = \mathcal{E}\{H\} < p + \epsilon.$$

Considérons alors la suite des nombres

$$N_0 = n_0, N_1 = N_0 + H_0, \dots, N_{k+1} = N_k + H_k, \dots;$$

quel que soit H_0 , ils comprennent la plupart des termes de la progression arithmétique

$$n_0, n_0 + p, n_0 + 2p, \dots, n_0 + \nu p, \dots,$$

la fréquence des termes qui manquent étant presque sûrement, à partir d'un certain moment, inférieure à $\epsilon/(p + \epsilon)$, donc à ϵ . On peut donc ne considérer que les valeurs de n de la forme N_k , et il suffit de démontrer que, parmi ces valeurs, la fréquence de celles pour lesquelles $\phi_n < x$ diffère de $F(x)$ d'au plus 2ϵ .

Nous allons montrer à cet effet que, ϕ'_k désignant la détermination de ϕ_n pour $n = N_k$, on peut appliquer la formule (8) pour $\phi = |\phi'_k|$ et $h = H_k$, et par suite aussi la formule (9), dans laquelle on peut évidemment identifier ψ_n avec la variable gaussienne désignée ci dessus par ψ'_k ; elle s'écrit donc

$$(9') \quad |\phi'_{k+1} - \psi'_k| \leq \epsilon(1 + |\psi'_k|).$$

Il n'y a qu'à procéder par récurrence. La définition de H_0 étant restée arbitraire, nous pouvons supposer ce nombre assez grand pour que le résultat énoncé soit vrai pour $k = 0$. Nous pouvons alors le supposer vrai pour une certaine valeur k . Il résulte dans ces conditions de la formule (9'), et de la définition de H (d'après laquelle $H \geq \Omega$), que l'on a

$$H_{k+1} \geq \frac{1}{|\log q|} \log \text{Max} \left(\frac{1}{\epsilon^2} \phi'^2_{k+1}, \frac{1}{2\epsilon - \epsilon^2} \right),$$

c'est-à-dire

$$\phi'_{k+1} q^{\frac{1}{2} H_{k+1}} \leq \epsilon, \quad 1 - q^{H_{k+1}} \geq (1 - \epsilon)^2.$$

La nouvelle application de la formule (8), et par suite celle de la formule (9') pour la valeur $k + 1$, sont donc justifiées.

La formule (10), conséquence des formules (8) et (9), s'applique donc aussi pour $n + h = N_k + H_k = N_{k+1}$, $\phi_{n+h} = \phi'_{k+1}$, $|\phi_n| = \phi = |\phi'_k|$. La fonction de répartition conditionnelle dont dépend ϕ'_{k+1} , lorsque ϕ'_k est connu diffère donc de $F(x)$ d'au plus ϵ , et, d'après la loi forte des grands nombres relative aux variables enchaînées, la fréquence des valeurs inférieures à x parmi les nombres $\phi'_1, \phi'_2, \dots, \phi'_k$ est presque sûrement, à partir d'une certaine valeur de k , comprise entre $F(x) - 2\epsilon$ et $F(x) + 2\epsilon$, c. q. f. d.

COROLLAIRE 1. *La fréquence des valeurs de n pour lesquelles on a*

$$(13) \quad \phi_n \leq x, \quad \phi_{n+1} \leq y$$

tend presque sûrement, pour n infini, vers la probabilité théorique de ces inégalités, déduites de la formule (4), c'est-à-dire vers l'expression

$$(14) \quad \frac{1}{2\pi\sqrt{1-q}} \iint_{\xi < x, \eta < y} \exp\left(-\frac{\xi^2 + \eta^2 - 2\xi\eta\sqrt{q}}{2(1-q)}\right) d\xi d\eta.$$

Indiquons seulement le principe de la démonstration, dont le lecteur reconstituera aisément les détails. Chaque petit intervalle de valeurs possibles pour ϕ_v est réalisé pour les valeurs $1, 2, \dots, n$ de v avec une fréquence peu différente, si n est grand, de sa probabilité théorique; donc un grand nombre de fois. On peut donc appliquer de nouveau la loi forte des grands nombres et conclure que ces valeurs de ϕ_v entraînent les différentes valeurs possibles pour ϕ_{v+1} avec des fréquences peu différentes de leur probabilité théorique.

Ce corollaire s'étend sans difficulté au cas où l'on considère simultanément un nombre quelconque de termes consécutifs de la suite des ϕ_n .

COROLLAIRE 2. *La fréquence des changements de signes dans la suite des ϕ_n tend presque sûrement vers $(1/\pi) \operatorname{Arc} \operatorname{tg} \sqrt{(1-q)/q}$ (donc vers $\frac{1}{4}$, si $q = \frac{1}{2}$). Ce corollaire est évidemment un cas particulier du corollaire 1.*

COROLLAIRE 3. *La fréquence des valeurs de n pour lesquels l'intervalle (t_{n+1}, t_n) contient au moins une racine de $X(t)$ tend presque sûrement vers $(2/\pi) \operatorname{Arc} \operatorname{tg} \sqrt{(1-q)/q}$.*

Ce corollaire résulte immédiatement du Corollaire 1, et de la loi forte des grands nombres. Si la suite des X_n est supposée connue, la fonction $X(t)$ devant être déterminée ensuite par des interpolations dans chacun des intervalles (t_{n+1}, t_n) , la fréquence des intervalles contenant au moins une racine est presque sûrement infiniment peu différente de la moyenne des probabilités conditionnelles théoriques. Rappelons que, pour chacun de ces intervalles, cette probabilité conditionnelle, évidemment égale à un pour $X_n X_{n+1} \leq 0$, est dans le cas contraire, d'après le lemme 2,

$$\exp\left(-\frac{2X_n X_{n+1}}{t_n - t_{n+1}}\right) = \exp\left(-\frac{2\phi_n \phi_{n+1}}{\sqrt{1-q}}\right).$$

Il résulte évidemment du corollaire 1 que cette probabilité conditionnelle est presque sûrement convergente en moyenne arithmétique vers sa valeur probable. La fréquence des intervalles (t_{n+1}, t_n) qui contiennent au moins une racine converge donc elle-même presque sûrement vers cette valeur prob-

able, c'est-à-dire vers la probabilité a priori de l'existence d'une racine dans l'intervalle (t_{n+1}, t_n) quand on ne connaît ni X_n , ni X_{n+1} . Cette probabilité a la valeur connue $(2/\pi) \text{Arc tg} \sqrt{(1-q)/q}$ [*Processus*, formules (42) et (44)^e], ce qui termine la démonstration du corollaire 3.

3. Le mouvement brownien plan. 1°. Ce mouvement est celui d'un point $A(t)$ dont les coordonnées rectangulaires $X(t)$ et $Y(t)$ sont deux fonctions aléatoires du type que nous venons d'étudier, indépendantes l'une de l'autre. Pour chaque valeur du paramètre t , que nous appellerons la *cote* de $A(t)$, ce point dépend de la loi de Gauss isotrope, d'écart type \sqrt{t} ; nous appelons ici écart type, non la valeur quadratique moyenne de la distance $R(t)$ du point $A(t)$ à l'origine 0, mais la valeur quadratique moyenne commune de $X(t)$ et $Y(t)$. On sait que $R(t)$ dépend de la loi défini par

$$(15) \quad \text{Pr}\{R(t) > r\} = e^{-r^2/2t}.$$

Les propriétés du vecteur $A(t)A(t+\tau)$, déplacement du point mobile pendant l'intervalle de temps $(t, t+\tau)$, sont naturellement indépendantes de t , et du passé; ce vecteur dépend de la loi de Gauss isotrope d'écart type $\sqrt{\tau}$.

De nombreuses propriétés du mouvement brownien plan sont des conséquences si évidentes des propriétés correspondantes du mouvement brownien linéaire qu'il suffit de les énoncer. Tel est le cas du principe d'interpolation: si $t_0 < t_1 < t_2$, et si $A(t_0)$ et $A(t_2)$ sont connus, la position probable de $A(t_1)$ est le point M_1 obtenu par une interpolation linéaire, et $M_1A(t_1)$ dépend de la loi de Gauss isotrope d'écart type

$$\sigma_1 = \sqrt{\frac{(t_1 - t_0)(t_2 - t_1)}{t_2 - t_0}}.$$

2°. Pour étudier la forme de la courbe au voisinage de l'origine 0, nous considérerons toujours la suite des valeurs $t_n = q^n$ de t ($0 < q < 1$). Nous écrirons A_n par abréviation de $A(t_n)$, et désignerons par M_n la position probable de ce point quand 0 et A_{n-1} sont connus; c'est le point défini par la formule vectorielle

$$\overline{0M_n} = q\overline{0A_{n-1}}.$$

On remarque que $\overline{0M_n}$ et $\overline{M_nA_n}$ sont deux vecteurs, indépendants l'un de l'autre, et dépendant respectivement des mêmes lois que $\overline{0A_{n+1}}$ et $\overline{A_{n+1}A_n}$; les propriétés stochastiques des deux triangles $0M_nA_n$ et $0A_{n+1}A_n$ sont donc identiques (si $q = \frac{1}{2}$, on peut écrire indifféremment $0M_nA_n$ ou A_nM_n0).

* Ces deux formules, que nous avons établies séparément, sont équivalentes; dans l'une, l'intervalle que nous désignons ici par (t_{n+1}, t_n) , est désigné par $(t, t+u)$; dans l'autre, il est désigné par $(t-u, t)$.

L'angle 0 de chacun de ces triangles, orienté et compté de $-\pi$ à $+\pi$, est une variable aléatoire θ dont la loi ne dépend que de q ; elle est symétrique; la valeur quadratique moyenne de θ est un nombre positif $\sigma = \sigma(q)$, inférieur à $\pi/2$.

3°. Nous désignerons par $R_n = \rho_n \sqrt{t_n}$ et Θ_n les coordonnées polaires de A_n ; nous prenons pour Θ_n la détermination obtenue en supposant $|\Theta_0| \leq \pi$ et en considérant les A_n comme des positions successives d'un point mobile qui se déplace sur la ligne polygonale $A_0 A_1 A_2 \dots$; chacun des accroissements $\Theta_{n+1} - \Theta_n = \theta_n$ est donc une variable aléatoire du type θ que nous venons de considérer.⁷ Mais les différents θ_n ne sont pas indépendants. Si l'on détermine successivement les points A_n dans l'ordre des n croissants, la loi conditionnelle dont dépend θ_n lorsque le passé est connu ne dépend que de R_n ; elle est toujours symétrique et comporte une valeur quadratique moyenne σ'_n fonction de R_n et q ; la valeur probable a priori de $\sigma'_n{}^2$ est naturellement σ^2 .

En ce qui concerne les R_n , on établit aisément le théorème suivant, analogue au théorème 3 relatif au mouvement brownien linéaire:

THÉORÈME 4. *La fréquence des valeurs supérieures à r dans la suite des ρ_n tend presque sûrement vers la limite*

$$Pr\{\rho_n > r\} = Pr\{R_n > r\sqrt{t_n}\} = e^{-r^2/2}.$$

Nous ne développerons pas la démonstration, tout à fait analogue à celle du théorème 3. On peut d'ailleurs aussi, en tenant compte de l'indépendance de $X(t)$ et de $Y(t)$, le considérer comme un corollaire du théorème 3.

Indiquons aussi, en ce qui concerne les R_n , la formule

$$(16) \quad Pr \left\{ \limsup_{n \rightarrow \infty} \frac{\rho_n}{\sqrt{2 \log n}} = 1 \right\} = 1,$$

analogue à la formule (5').

Comme il est évident que $R_n \geq |X_n|$, donc $\rho_n \geq |\phi_n|$, la borne inférieure donnée pour $|\phi_n|$ par la formule (5') s'applique à ρ_n . Pour établir la formule (16), il reste à montrer que, pour $c > 1$, il existe presque sûrement un N tel que, pour $n > N$, on ait

$$(17) \quad \rho_n < c\sqrt{2 \log n}.$$

Cela résulte évidemment de ce que la probabilité de l'inégalité inverse, n^{-c^2} , est le terme général d'une série convergente.

⁷ Il y a indétermination dans la valeur à prendre pour Θ_n si l'un des $|\theta_v|$ ($v = 0, 1, \dots, n-1$) a la valeur π , c'est-à-dire si la ligne $A_0 A_1 \dots A_n$ passe en 0. La probabilité de cette circonstance étant nulle, il n'en résulte aucune difficulté.

On peut aussi déduire la formule (17) de la formule (6), et de la remarque que l'angle Θ_0 , dont dépend l'orientation de la courbe autour de l'origine, est choisi au hasard et indépendant de la suite des ρ_n . Si alors on pouvait trouver, avec une probabilité positive, des valeurs de n arbitrairement grandes pour les quelles on ait $\rho_n \geq c\sqrt{2 \log n}$, il en serait de même en imposant la condition supplémentaire $\cos \Theta_n > \cos \alpha$, α étant assez petit pour que $c \cos \alpha = c' > 1$; cela est en contradiction avec la formule (6), écrite en remplaçant c par c' .

Les grandes valeurs des ρ_n ne sont donc pas plus grandes que celles des $|\phi_n|$; dans un cas comme dans l'autre, on peut appeler grandes valeurs celles qui sont supérieures à $(1 - \epsilon)\sqrt{2 \log n}$, ϵ étant un nombre positif donné et très petit. Mais bien entendu les grandes valeurs de ρ_n sont plus fréquentes que celles des $|\phi_n|$. Elles correspondent à des vecteurs OA_n ayant, à un angle arbitrairement petit près, toutes les orientations possibles, et il faut choisir ceux qui font avec une direction donnée un angle très petit ou très voisin de π pour trouver une grande valeur de $|\phi_n|$.

4°. Occupons-nous maintenant des angles θ_n et Θ_n . Remarquons d'abord que, comme conséquence du fait que les différentes valeurs possibles pour ρ_n sont réalisées avec des fréquences tendant vers leurs probabilités théoriques, et de ce que la nature stochastique du triangle $OA_n A_{n+1}$ est fonction de ρ_n supposé connu, les différentes formes possibles pour ce triangle, et par suite les différentes valeurs de θ_n , sont aussi presque sûrement réalisées avec des fréquences tendant vers leurs probabilités théoriques. Il s'agit d'une nouvelle application du principe utilisé pour le corollaire 1 du théorème 3.

Notons surtout que, σ'_n étant une fonction (non aléatoire, bornée, et continue) de ρ_n , les différentes valeurs de σ'_n sont réalisées avec des fréquences tendant presque sûrement vers leurs probabilités théoriques, et l'on a presque sûrement

$$(18) \quad \lim_{n \rightarrow \infty} \frac{\sigma'^2_1 + \sigma'^2_2 + \dots + \sigma'^2_n}{n} = \mathcal{E}\{\sigma_v^2\} = \sigma^2,$$

et par suite aussi

$$(18') \quad \lim_{n \rightarrow \infty} \frac{1}{n} (\theta_1^2 + \theta_2^2 + \dots + \theta_n^2) = \sigma^2.^8$$

⁸ Cette formule peut être soit obtenue comme application directe du premier alinéa du présent § 3, 4°, soit déduite de la formule (18) et d'une formule que nous avons établie antérieurement [Var. aléatoires, formule (22), p. 252] d'après laquelle il y a presque sûrement convergence en moyenne arithmétique vers zéro de la suite des variables $\theta_n^2 - \sigma'^2_n$.

Supposons que l'on détermine d'abord les modules, puis les signes des θ_v . Après détermination des modules, et en supposant OA_0 pris comme origine des angles polaires, Θ_n se présente sous la forme

$$(19) \quad \Theta_n = \epsilon_1 |\theta_1| + \epsilon_2 |\theta_2| + \dots + \epsilon_n |\theta_n|,$$

les signes ϵ_v étant choisis au hasard indépendamment les uns des autres. Comme les $|\theta_v|$ sont bornés, et que la série $\sum \theta_v^2$ est divergente, il résulte du second théorème limite du calcul des probabilités que Θ_n est asymptotiquement une variable gaussienne. Son écart type, d'après (18'), est un infiniment grand équivalent à $\sigma\sqrt{n}$. Donc $\Theta_n/\sigma\sqrt{n}$ est asymptotiquement une variable gaussienne réduite. En termes précis: ϵ étant un nombre positif arbitrairement petit, il existe presque sûrement un nombre N tel que, pour x quelconque et $n > N$, on ait

$$(20) \quad |Pr'\{\Theta_n < \sigma x \sqrt{n}\} - F(x)| < \epsilon,$$

Pr' désignant une probabilité conditionnelle, évaluée en supposant connus les $|\theta_v|$.

Le nombre N est aléatoire; mais, en négligeant des cas de probabilité inférieure à ϵ , on peut lui assigner une borne supérieure non aléatoire n' . Comme Pr est la valeur probable de Pr' , et que, dans les cas où l'inégalité (20) n'est pas vérifiée, son premier membre est du moins au plus égal à un, on a, pour $n > n'$

$$(21) \quad |Pr\{\Theta_n < \sigma x \sqrt{n}\} - F(x)| < 2\epsilon,$$

c'est-à-dire que: Θ_n est asymptotiquement une variable gaussienne; son écart type est un infiniment grand équivalent à $\sigma\sqrt{n}$.

L'angle Θ_n apparaît ainsi comme le gain d'un joueur dans une partie de pile ou face à enjeu aléatoire, pouvant dépendre des enjeux antérieurs, et déterminé pour chaque coup par une expérience préalable. La formule presque sûre (18) permet d'assimiler cette partie à une partie à enjeu fixe σ .⁹

La plupart des résultats connus relatifs à une telle partie s'appliquent de même ici, notamment la loi du logarithme itéré, d'après laquelle on a presque sûrement

$$(22) \quad \limsup_{n \rightarrow \infty} \frac{\Theta_n}{\sqrt{2n \log \log n}} = \sigma,$$

le même résultat s'appliquant à $-\Theta_n$. La ligne polygonale $A_1 A_2 \dots A_n \dots$ tourne donc indéfiniment (et fort irrégulièrement) autour du point 0 avant de l'atteindre, Θ_n ayant des valeurs arbitrairement grandes des deux signes.

⁹ On peut aussi appliquer directement le second théorème limite du calcul des probabilités sous la forme, applicable à certaines suites de variables enchaînées, que

5°. Désignons par Θ'_p l'angle $A_0 O A_p$, compté de $-\pi$ à $+\pi$. Il diffère de Θ_p par un multiple de 2π qui, si $p > 1$, a une probabilité positive de n'être pas nul. Or, dans ce cas, $|\Theta'_p| < |\Theta_p|$. La valeur quadratique moyenne de Θ'_p est donc inférieure à celle de Θ_p , ce qui s'exprime par la formule

$$(23) \quad \sigma(q^p) < \sqrt{p} \sigma(q).$$

Cette formule permet de comparer les angles polaires obtenus pour un même point A_{np} si l'on va de A_0 à ce point, d'une part en suivant le ligne polygonale $A_0 A_1 A_2 \dots A_{np}$, d'autre part en suivant la ligne raccourcie $A_0 A_p A_{2p} \dots A_{np}$. Les valeurs quadratiques moyennes de ces angles sont respectivement $\sqrt{np} \sigma(q)$ et $\sqrt{n} \sigma(q^p)$. D'après la formule (23), la seconde est plus petite; cela était à prévoir: la seconde ligne évite les détours de la première.

Inversement, on peut suivre de plus en plus exactement les détours de la courbe C en prenant pour q des valeurs de plus en plus voisines de l'unité. Pour que A_n coïncide avec le point $A(t)$ correspondant à une valeur de t donnée entre zéro et un, nous prendrons $q = t^{1/n}$; la valeur quadratique moyenne de l'angle polaire obtenu pour $A(t)$ est alors $\sqrt{n} \sigma(q)$. Il y a lieu de s'attendre à ce qu'elle croisse avec n , puisqu'en prenant pour n des valeurs de plus en plus grandes on suit de mieux en mieux les détours de la courbe; on est sûr, par ce qui précède, qu'elle croît quand on passe d'une valeur initiale n_0 à une valeur n_1 multiple de n_0 . Il y a intérêt, pour connaître la nature de l'angle polaire $\Theta(t)$ obtenu en suivant la courbe elle-même (angle bien défini; il s'agit d'une courbe continue ne passant en général pas par l'origine, comme nous le verrons plus loin), de chercher à définir l'expression

$$(24) \quad \mathcal{E}\{\Theta^2(t)\} = \lim_{n \rightarrow \infty} n \sigma^2(t^{1/n}) = \log \frac{1}{t} \lim_{q \rightarrow 1} \frac{\sigma^2(q)}{1-q}.$$

Nous allons montrer que cette expression est infinie.

A cet effet, H_1 désignant le pied de la perpendiculaire abaissée de A_1 sur OA_0 , et ρ et r étant des nombres positifs, considérons l'hypothèse

$$(E) \quad OA_0 > \rho, \quad |A_0 H_1| < r\sqrt{1-q}, \quad |H_1 A_1| < r\sqrt{1-q}.$$

Si elle est réalisée, quand q tend vers 1, $A_0 A_1$ est petit, et $A_0 O A_1$ est un infiniment petit équivalent (au sens de Bernoulli) à $H_1 A_1 / OA_0$. On en déduit

$$(25) \quad \liminf_{q \rightarrow 1} \frac{\sigma^2(q)}{1-q} \geq \text{Pr}\{E\} \mathcal{E}' \left\{ \frac{H_1 A_1^2}{1-q} \right\} \mathcal{E}' \left\{ \frac{1}{OA_0^2} \right\};$$

j'ai indiquée antérieurement (Var. aléatoires, pp. 237-242). J'ai préféré ici profiter de la symétrie des lois dont dépendent les θ_n , et de l'indépendance des ϵ_n , pour indiquer une démonstration plus simple.

dans cette formule, \mathcal{E}' désigne une valeur probable calculée dans l'hypothèse E ; en ne tenant compte que des cas où cette hypothèse est vérifiée, on a bien une borne inférieure de la valeur probable de $\theta_1^2/(1-q)$ qu'il s'agit d'évaluer. Si maintenant ρ et $1/r$ tendent vers zéro, les deux premiers facteurs tendent vers l'unité; le dernier facteur

$$e^{\rho^2/2} \int_{\rho}^{\infty} e^{-u^2/2} du/u$$

augmente indéfiniment; il en est donc de même de $\sigma^2(q)/(1-q)$, c. q. f. d.

On s'explique aisément ce résultat en observant que, si la courbe passe très près de l'origine, l'angle polaire varie rapidement. Or, sans avoir une probabilité positive de passer exactement à l'origine, l'arc $A(t)A(1)$ a une probabilité positive d'en passer arbitrairement près; on s'explique aisément que cette probabilité soit suffisante pour que la valeur quadratique moyenne de $\Theta(t)$ soit infinie.

Revenant alors au cas où q est fixe, considérons la suite des points A_n , et les valeurs qui leur correspondent de l'angle polaire $\Omega_n = \Theta(t_n)$. Cet angle apparaît comme une somme de termes aléatoires indépendants dépendant d'une même loi symétrique et à valeur quadratique moyenne infinie. On sait que dans ces conditions, par l'effet de ces grandes valeurs qui se trouvent réalisées de temps en temps quand n augmente, l'ordre de grandeur à prévoir pour Ω_n est, si n est grand, supérieur à celui de \sqrt{n} . Ces grandes valeurs de temps en temps réalisées correspondant à des arcs $A_n A_{n+1}$ qui passent très près de l'origine, on voit que, si OA_n est en général de l'ordre de grandeur de $\sqrt{t_n}$, il y a parfois des valeurs plus petites [et aussi des valeurs plus grandes, comme le montre la formule (16)]. Il en résulte que, quand t tend vers zéro, $OA(t)$, qui doit finalement devenir nul, varie fort irrégulièrement; la courbe a l'aspect d'une succession de boucles qui se ferment de plus en plus près de l'origine.

Il peut être intéressant de préciser d'avantage. Disons seulement que la loi à valeur quadratique moyenne infinie que nous venons de considérer a, pour toute exposant $\alpha < 2$, une moyenne d'ordre α finie; elle appartient au domaine d'attraction de la loi de Gauss. Par suite la variation de l'angle polaire Θ sur un arc $A(t')A(t'')$, divisée par une fonction convenable de t'/t'' , dépend d'une loi qui tend vers la loi de Gauss réduite quand ce rapport augmente indéfiniment ou tend vers zéro.

6°. Les résultats qui précèdent s'appliquent évidemment à l'étude de la courbe C au voisinage du point correspondant à n'importe quelle valeur donnée de t , soit à gauche, soit à droite de ce point (*gauche* signifiant ici du côté des t décroissants; *droite*, du côté des t croissants). Nous appellerons *tangente*

en un point une droite passant par ce point et laissant d'un même côté un petit arc de courbe contenant ce point; il peut y avoir des *demi-tangentes*, à gauche, ou à droite. Au point $A(t)$ correspondant à une valeur de t donnée, ou choisie au hasard par une expérience indépendante du choix de C , il n'y a presque sûrement ni tangente, ni demi-tangente.

Par contre, il existe presque sûrement des points exceptionnels, où il y a des tangentes. Nous avons déjà observé que tout intervalle de variation de t contient une infinité dénombrable de maxima et minima de $X(t)$; chacun d'eux correspond à une tangente parallèle à l'axe des y . Ce résultat s'appliquant aux tangentes parallèles à n'importe quelle direction, il y a une infinité continue de tangentes. Il y a d'autre part une double infinité de demi-tangentes: toute droite coupant la courbe est une demi-tangente à chacune des extrémités des intervalles extérieurs à la courbe.

Il n'y a presque sûrement aucune tangente double parallèle à l'axe des y ; les maxima et minima de $X(t)$ constituent en effet une infinité dénombrable, de sorte que l'ensemble des valeurs obtenues pour ces maxima et minima n'a aucune chance de contenir une valeur qui soit obtenue une seconde fois. La même remarque s'applique aux tangentes doubles parallèles à une direction donnée. Par contre, comme nous allons le montrer, *il y a presque sûrement une infinité dénombrable de tangentes doubles* (mais la probabilité que l'une d'elles ait une direction donnée d'avance est nulle).

Considérons à cet effet deux arcs $A(t_0')A(t_1')$ et $A(t_0'')A(t_1'')$ de C , et désignons respectivement par Γ' et Γ'' les plus petits contours convexes entourant respectivement ces arcs. Il existe de zéro à quatre tangentes communes à Γ' et Γ'' ; comme $A(t_0')$ et $A(t_1')$ sont presque sûrement intérieurs à Γ' , et que $A(t_0'')$ et $A(t_1'')$ sont intérieurs à Γ'' , si t_0' , t_1' , t_0'' et t_1'' sont choisis au hasard, ces droites sont presque sûrement des tangentes doubles à C .

Sur n'importe quel arc de C , nous pouvons choisir deux points distincts $A(t')$ et $A(t'')$, puis prendre $t_0' - t'$, $t_1' - t'$, $t_0'' - t''$, $t_1'' - t''$, assez petits pour que Γ' et Γ'' soient extérieurs l'un à l'autre et que les droites que nous venons de considérer existent. Pour tout arc de C , il y a donc des tangentes doubles; donc il y en a une infinité dénombrable au moins.

D'autre part, pour toute tangente double $A(t')A(t'')$, on peut définir dans l'espace représentant l'ensemble des quatre nombres t_0' , t_1' , t_0'' , t_1'' , un domaine $(t' - t_0'$, $t_1' - t'$, $t'' - t_0''$, $t_1'' - t''$ positifs et assez petits) tel que, pour tout point de ce domaine, $A(t')A(t'')$ soit une des quatre tangentes doubles (au plus) que l'on peut définir en partant de ce point. Il ne peut donc y avoir qu'une infinité dénombrable de tangentes doubles.

Désignons par Γ le plus petit contour convexe entourant un arc

$A(t')A(t'')$. Nous allons montrer qu'il est presque sûrement constitué par une infinité dénombrable de tangentes doubles, formant un ensemble partout dense autour de Γ , c'est-à-dire que n'importe quelle tangente à Γ est limite de tangentes doubles; il n'y a pas de point anguleux. En d'autres termes, si M est le point de Γ d'abscisse curviligne s , la tangente en M à Γ fait avec une direction fixe un angle θ qui est une fonction de s continue, évidemment monotone, et à dérivée presque partout nulle.

La démonstration va résulter de propriétés presque sûres de C et de Γ au voisinage du point de Γ où x est minimum: la courbure est infinie, mais θ varie d'une manière continue. Ces propriétés doivent de même être vérifiées au point de contact de la tangente définie par une valeur de θ choisie au hasard; elles ne peuvent donc être en défaut que pour des valeurs de θ ayant une probabilité nulle d'être choisies, c'est-à-dire constituant un ensemble de mesure nulle. Il n'en serait pas ainsi s'il y avait sur Γ des points anguleux, ou si l'ensemble des points pour lesquels la courbure est positive et finie constituait sur Γ un ensemble de mesure linéaire positive.

Considérons donc le minimum de x , que nous pouvons évidemment supposer réalisé à l'origine, et pour $t = 0$. Il suffit aussi de considérer l'arc voisin de ce point correspondant aux valeurs positives de t . Nous avons étudié antérieurement les propriétés $X(t)$ dans ces conditions (Processus, § 9, 1° et 2°): $X(t)$, pour une valeur de t très petite et choisie au hasard, est en général de l'ordre de grandeur de \sqrt{t} ; les grandes valeurs, fortuitement, mais presque sûrement réalisées, sont de l'ordre de grandeur de $\sqrt{2t \log |\log t|}$; les petites valeurs sont de l'ordre de grandeur de $\sqrt{t} |\log t|^{-1+\epsilon}$, ϵ tendant vers zéro avec t .

D'autre part la variation de $Y(t)$ est indépendante de l'hypothèse que $X(t)$ soit positif. Il est donc presque sûr que, dans la suite des valeurs de t tendant vers zéro et correspondant soit aux petites valeurs, soit aux grandes valeurs, de $X(t)$, on trouve des valeurs arbitrairement grandes de $Y(t)/\sqrt{t}$ (positives ou négatives), mais que $Y(t)/\sqrt{2t \log |\log t|}$ est borné (et asymptotiquement ≤ 1).

De la comparaison des inégalités ainsi obtenues pour $X(t)$ et $Y(t)$ résulte d'une part que $Y(t)/X(t)$ prend toutes les valeurs possibles entre $-\infty$ et $+\infty$, d'autre part que $Y^2(t)/X(t)$ tend vers zéro. L'hypothèse d'une courbure finie et celle d'un point anguleux sont ainsi exclues, c. q. f. d.

Indiquons enfin sans démonstration l'extension suivante des résultats précédents: dans le mouvement brownien à p dimensions, il est presque sûr que, pour n'importe quel arc de trajectoire, la plus petite hypersurface convexe qui le contienne à son intérieur a un plan tangent bien défini en tout point et

qui varie d'une manière continue, mais ne comporte aucune partie courbe et est constituée par une infinité dénombrable de faces planes. On peut donc, dans l'espace considéré, faire varier l'orientation d'une variété linéaire à deux dimensions sans que pour aucune d'elles le mouvement projeté sur cette variété mette en défaut les propriétés du mouvement brownien que nous venons d'obtenir. Ce résultat va beaucoup plus loin que celui qui consiste à dire que, pour chaque arc de chaque courbe C considéré isolément, elles ont une probabilité égale à l'unité.

4. La notion d'oscillation brownienne. 1°. Nous nous placerons d'abord dans le cas du mouvement linéaire, et supposons que t varie de zéro à un. Supposons cet intervalle divisé en un grand nombre n d'intervalles partiels dont le plus grand ait une longueur très petite ϵ_n ; désignons par Δt un quelconque de ces intervalles, par ΔX la variation correspondante de $X(t)$, et posons

$$(26) \quad b_n = \sum (\Delta t)^2, \quad \beta_n = \sum (\Delta X)^2.$$

On a évidemment

$$\mathcal{E}\{B_n\} = \sum \Delta t = 1$$

$$\mathcal{E}\{(B_n - 1)^2\} = \sum \mathcal{E}\{[(\Delta X)^2 - \Delta t]^2\} = 3b_n - 2b_n + b_n$$

et par suite

$$(27) \quad \mathcal{E}\{(B_n - 1)^2\} = 2b_n \leq 2 \sum \Delta t \text{ Max } \Delta t = 2\epsilon_n.$$

Si donc on fait varier le mode de division de l'intervalle $(0, 1)$ en intervalles partiels de manière que ϵ_n tende vers zéro, il y a convergence en moyenne quadratique de B_n vers l'unité; donc aussi convergence en probabilité.

Nous allons compléter ce résultat par l'étude de cas où il y a convergence presque sûre.

Désignons par B'_p la valeur de B_n lorsque l'intervalle $(0, 1)$ est divisé en 2^p intervalles égaux et montrons d'abord que: B'_p tend presque sûrement vers l'unité.

C'est une conséquence immédiate de la formule (27), qui s'écrit, dans le cas considéré

$$\mathcal{E}\{(B'_p - 1)^2\} = 2^{1-p}.$$

L'inégalité de Tchebycheff donne alors

$$Pr\{|B'_p - 1| \geq p/2^{p/2}\} \leq 2/p^2,$$

et, cette expression étant le terme général d'une série convergente, il existe presque sûrement une valeur de p à partir de laquelle on a

$$|B'_p - 1| < p/2^{p/2},$$

ce qui démontre et précise le résultat annoncé.

Il serait facile, en ne faisant que des raisonnements très simples, de généraliser ce résultat. Nous allons établir un théorème plus général, qui comprend toutes ces généralisations presque évidentes.

2°. Considérons une suite de valeurs $t_1, t_2, \dots, t_v, \dots$ de t , comprises entre zéro et un, et formant un ensemble partout dense dans l'intervalle $(0, 1)$. Les $n - 1$ premiers nombres t_v définissent une division de cet intervalle en n intervalles partiels, à laquelle nous associerons comme tout à l'heure la somme non aléatoire b_n et la somme aléatoire B_n ; b_n , au plus égal au plus grand des intervalles partiels, tend vers zéro pour n infini.

THÉORÈME 5. On a

$$Pr\{\lim_{n \rightarrow \infty} B_n = 1\} = 1.$$

Pour le démontrer, observons que, quand n augmente d'une unité, la variation de la somme b_n provient d'un seul terme, que nous désignerons par τ_n^2 , qui se trouve remplacé par une somme $\tau_n'^2 + \tau_n''^2 = \tau_n^2 - 2\tau_n'\tau_n''$. On en déduit

$$(28) \quad b_n - b_{n+1} = 2\tau_n'\tau_n'',$$

et, pour B_n , on a de même

$$(29) \quad B_n - B_{n+1} = 2\xi_n'\xi_n'',$$

ξ_n' et ξ_n'' étant les accroissements de $X(t)$ dans deux intervalles contigus, de longueurs respectives τ_n' et τ_n'' . Comme ils sont indépendants, on a

$$(30) \quad \mathcal{E}\{\xi_n'\xi_n''\} = 0, \quad \mathcal{E}\{\xi_n'^2\xi_n''^2\} = \tau_n'\tau_n''.$$

Proposons nous maintenant d'étudier l'oscillation de B_n quand n varie dans un intervalle (p, q) ; nous poserons

$$T_{p,q} = \max_{p \leq n \leq q} |B_n - B_q|.$$

Nous considérerons d'abord des probabilités, que nous désignerons par des lettres accentuées (Pr' ou \mathcal{E}'), évaluées en supposant connus les termes de B_q , c'est-à-dire que l'on connaît les valeurs absolues des accroissements ΔX dont les carrés interviennent dans B_q , mais non leurs signes. D'après (29), les déterminations successives de $B_{q-1}, B_{q-2}, \dots, B_p$ dépendent des signes de $\xi_n'\xi_n''$ pour les valeurs $q - 1, q - 2, \dots, p$ de n . Ces signes sont indépendants; pour tout n inférieur à q , $|\xi_n'|$ et $|\xi_n''|$ étant connus, les deux signes possibles pour $\xi_n'\xi_n''$ sont également probables; le choix d'un signe détermine $|\xi_n| = |\xi_n' + \xi_n''|$, et l'on se retrouve dans les mêmes conditions pour déterminer $|\xi_{n-1}|$ par un nouveau groupement de termes.

Nous sommes ainsi dans les conditions voulues pour appliquer une

inégalité connue de A. Kolmogoroff, ou du moins son extension au cas de certaines sommes de variables enchaînées que nous avons indiquée antérieurement (*Var. aléatoires*, pp. 246-247), et il vient

$$Pr\{T_{p,q} \geq \epsilon\} \leq \frac{4}{\epsilon^2} \sum_{p \leq n < q} \mathcal{E}'\{\xi_n'^2 \xi_n''^2\}.$$

Où Pr et \mathcal{E} sont respectivement les valeurs probables de Pr' et \mathcal{E}' . Compte tenu des formules (28) et (30), il vient

$$(31) \quad Pr\{T_{p,q} \geq \epsilon\} \leq \frac{2}{\epsilon^2} (b_p - b_q) < \frac{2}{\epsilon^2} b_p.$$

Posons maintenant

$$T_p = \lim_{q \rightarrow \infty} T_{p,q}, \quad T = \lim_{p \rightarrow \infty} T_p.$$

Ces limites, finies ou infinies, existent presque sûrement, à cause du caractère monotone de $T_{p,q}$, et, b_p tendant vers zéro pour p infini, on déduit de l'inégalité (31)

$$(32) \quad Pr\{T_p \geq \epsilon\} \leq \frac{2}{\epsilon^2} b_p, \quad Pr\{T \geq \epsilon\} = 0.$$

Comme cela est vrai quelque petit que soit ϵ , il est presque sûr que $T = 0$, c'est-à-dire que B_n a une limite B . Comme enfin il y a convergence en probabilité vers l'unité, on a $B = 1$, c. q. f. d.

Naturellement, comme dans tous les énoncés de cette nature, il y a convergence uniforme de la suite des B_n , sauf dans des cas de probabilité inférieure à un nombre arbitrairement petit η . On peut en effet, d'après (32), déterminer p_h (pour $h = 1, 2, \dots$) de manière que

$$Pr\{T_{p_h} \geq \frac{1}{h}\} \leq 2h^2 b_{p_h} < \frac{6\eta}{\pi^2 h^2},$$

et par suite

$$Pr\{T_{p_1} < 1, T_{p_2} < \frac{1}{2}, \dots, T_{p_h} < \frac{1}{h}, \dots\} > 1 - \frac{6\eta}{\pi^2} \sum \frac{1}{h^2} = 1 - \eta.$$

Comme, dans les cas de convergence vers l'unité, $n' \geq n$ entraîne $|B_{n'} - 1| \leq T_n$, le résultat énoncé est bien établi.

On remarque aussi que la convergence obtenue est indépendante du choix des t_n , si l'on assujettit ce choix à la seule condition que ϵ_n (donc aussi b_n) soit borné supérieurement par une fonction donnée de n qui tende vers zéro pour n infini.

3°. Introduisons maintenant le hasard dans le choix des t_n . Nous supposons ces nombres choisis successivement d'après des lois qui peuvent n'être

pas indépendantes les unes des autres ; mais, pour chaque n , la loi à n variables t_1, t_2, \dots, t_n est bien déterminée et indépendante des expériences qui déterminent $X(t)$; de plus ces lois doivent être telles que ϵ_n tende presque sûrement vers zéro. Tel sera le cas si, par exemple, pour n'importe quel intervalle Δt , la probabilité que t_n soit dans cet intervalle a une borne inférieure indépendante de t_1, t_2, \dots, t_{n-1} , et qui soit le terme général d'une série divergente.

La suite des B_n dépend alors de deux séries d'expériences indépendantes l'une de l'autre. La première a pour objet de déterminer la suite des t_n , et l'on sait (cf. *Var. aléatoires*, p. 22) qu'il est en tout cas possible de faire correspondre les différentes suites possibles aux différentes valeurs d'une variable T comprise entre zéro et un, et cela de manière que la probabilité de n'importe quel ensemble de suites possibles soit égale à la mesure de l'ensemble des valeurs de T qui leur correspondent (si cet ensemble n'est pas mesurable, la probabilité est indéterminée entre une *probabilité intérieure* et une *probabilité extérieure*, égales respectivement à la mesure intérieure et à la mesure extérieure de cet ensemble) ; chacun des t_n est une fonction mesurable de T . D'une manière analogue, on peut représenter par une variable unique U l'ensemble des choix qui déterminent successivement $X(1), X(\frac{1}{2}), X(\frac{1}{4}), X(\frac{3}{4}), \dots$, et par suite toutes les fonctions $X_n(t)$ du théorème 1 ; $X(t)$ est la limite presque sûre de $X_n(t)$, la convergence étant uniforme (en t et U) en dehors d'un ensemble de valeurs de U de mesure arbitrairement petite.

Désignons par E l'ensemble des points T, U , du carré $0 \leq T \leq 1, 0 \leq U \leq 1$, pour lesquels on ait $\lim B_n = 1$; par E' l'ensemble complémentaire. Nous allons montrer que

THÉORÈME 6. *L'ensemble E' est mesurable et de mesure nulle.*

Si l'on admet que E est mesurable, la démonstration est immédiate : il est presque sûr que $\lim \epsilon_n = 0$, et que cela entraîne $\lim B_n = 1$. En d'autres termes, sauf pour des valeurs de t constituant un ensemble de mesure nulle, l'ensemble des points de E' situé sur la droite $T = t$ a une mesure linéaire nulle. Donc E' a une mesure superficielle nulle.

Il reste à montrer que E est mesurable. On sait que la probabilité de la convergence d'une suite de variables aléatoires, et celle de sa convergence vers une limite donnée, sont toujours bien déterminées. Pour montrer que ce théorème est ici applicable, il faut montrer, non seulement que chaque inégalité $B_n < \beta$ a une probabilité déterminée, c'est-à-dire que B_n est une fonction mesurable du point T, U , mais qu'il en est de même de toute combinaison en nombre fini d'inégalités de cette forme. Ce second résultat est d'ailleurs une conséquence évidente, non du premier résultat considéré isolément, mais de ce résultat et du fait qu'une même représentation du résultat des expériences sur

le plan des T, U permet d'étudier toutes les fonctions B_n ; on sait en effet que, dans ce plan, la partie commune à plusieurs ensembles mesurables est un ensemble mesurable. On est donc ramené à démontrer que chaque fonction B_n est une fonction mesurable des point T, U .

Nous démontrerons un résultat plus général, qui aura plus loin une autre application: si $\phi(x_1, x_2, \dots, x_n)$ est une fonction continue de l'ensemble de ses n arguments, l'expression

$$(33) \quad \Phi = \phi[X(t_1), X(t_2), \dots, X(t_n)]$$

est une fonction mesurable du point T, U .

La démonstration est immédiate, en utilisant la définition de $X(t)$ comme limite des approximations $X_v(t)$. En remplaçant $X(t)$ par $X_v(t)$, Φ se trouve remplacé par une expression Φ_v qui est une fonction continue de t_1, t_2, \dots, t_n , et des 2^v quantités $X(h/2^v)$ qui interviennent dans la détermination de $X_v(t)$. C'est une fonction continue d'un nombre fini de fonctions mesurables de T ou de U , donc du point T, U ; c'est donc une fonction mesurable de ce point.

Il suffit donc de montrer que Φ_v tend en mesure vers Φ , pour v infini; c'est-à-dire que, ϵ et ϵ' étant arbitrairement petits, on peut déterminer v' tel que, pour $v > v'$, on ait

$$Pr\{|\Phi - \Phi_v| > \epsilon'\} < \epsilon.$$

Or on peut d'abord déterminer M tel que

$$Pr\{\text{Max } |X(t)| > M\} < \frac{\epsilon}{2}.$$

Les nombres $X(t_1), X(t_2), \dots, X(t_n), X_v(t_1), X_v(t_2), \dots, X_v(t_n)$ étant ainsi bornés (en dehors de cas de probabilité inférieure à $\epsilon/2$), on n'a à considérer qu'une région où la fonction $\phi(x_1, x_2, \dots, x_n)$ est uniformément continue: si donc chacun des $X(t_h) - X_v(t_h)$ ($h = 1, 2, \dots, n$) ne dépasse pas en valeur absolue un certain module de continuité $\eta = \eta(\epsilon')$, on a $|\Phi - \Phi_v| \leq \epsilon'$. Or nous avons vu qu'en négligeant des cas de probabilité inférieure à un nombre arbitrairement petit (nous prendrons ici $\epsilon/2$), $|X(t) - X_v(t)|$ peut, pour tout t entre zéro et un et tous les cas non négligés, être rendu inférieur à un nombre arbitrairement petit (ici η); il suffit que v soit assez grand. Dans ces conditions on a bien $|\Phi - \Phi_v| \leq \epsilon'$, sauf dans les cas négligés dont la probabilité totale est inférieure à $\epsilon/2 + \epsilon/2 = \epsilon$, c. q. f. d.

COROLLAIRE. La partie $B_n(t)$ de la somme B_n qui dépend des valeurs de $X(u)$ dans l'intervalle $(0, t)$ tend presque sûrement, pour n infini, vers $B(t) = t$ et cela uniformément quand t varie de zéro à un.

Le théorème précédent s'applique évidemment pour chaque valeur de t comme pour la valeur un.¹⁰ En considérant alors un ensemble dénombrable de valeurs t'_ν de t , partout dense entre zéro et un, il y a convergence presque sûre de chacun des $B_n(t'_\nu)$ vers $B(t'_\nu)$; en effet, pour chaque t'_ν , il n'y a divergence que dans des cas dont la probabilité est nulle; la réunion de tous ces cas a encore une probabilité nulle. En dehors de ces cas, il y a en tous les points t'_ν convergence de la fonction monotone $B_n(t)$ vers la limite $B(t)$ continue, donc uniformément continue, dans l'intervalle fermé $(0, 1)$. On sait qu'il y a alors convergence uniforme dans tout l'intervalle, c. q. f. d.

4°. Arrivons maintenant à une application du théorème de Fubini, qui est fondamentale. Désignons par F l'ensemble du plan des T, U , intérieur au carré $0 \leq T \leq 1, 0 \leq U \leq 1$, et correspondant à l'ensemble des cas où il y a convergence uniforme de $B_n(t)$ vers t dans l'intervalle $(0, 1)$. On peut indifféremment calculer sa mesure en intégrant par rapport à T la mesure linéaire de sa section par une droite $U = \text{const.}$, ou en faisant l'inverse. C'est par la première méthode de calcul que nous avons déterminé cette mesure, et montré que le complément de F est de mesure nulle. L'interversion de l'ordre des intégrations nous donne immédiatement un résultat important. Pour l'énoncer simplement, nous dirons qu'une fonction $X(t)$ est un modèle de mouvement brownien linéaire si, la suite des t_n étant choisie au hasard, on obtient avec une probabilité unité une suite de fonctions $B_n(t)$ ayant une limite non aléatoire $B(t)$, fonction continue et croissante de t ; pour chaque intervalle Δt , la variation $\Delta B(t)$ sera la mesure de l'oscillation brownienne.

La conséquence annoncée du fait que l'ensemble F ait pour mesure l'unité s'énonce alors ainsi:

THÉORÈME 7. *Le schéma stochastique du mouvement brownien linéaire réalise avec une probabilité unité un modèle de mouvement brownien linéaire; de plus $B(t) = t$.*

Quelques remarques sont nécessaires pour bien comprendre la définition qui précède. Il est d'abord évident que, pour une fonction donnée $X(t)$, il peut arriver que $B_n(t)$ ait une limite presque sûre autre que t . Cette limite est nécessairement une fonction non décroissante de t . Si elle est constante dans un intervalle, c'est que la fonction $X(t)$ n'y est pas assez irrégulière pour pouvoir donner une idée du mouvement brownien. Il est peut-être aussi possible, si $X(t)$ a au voisinage d'un point une allure trop irrégulière, que

¹⁰ On remarque d'ailleurs qu'en raison de l'indépendance des oscillations de $X(u)$ dans les deux intervalles $(0, t)$ et $(t, 1)$, il ne peut avoir convergence presque sûre de $B_n = B_n(1)$ vers une limite que si $B_n(t)$ et $B_n - B_n(t)$ ont séparément des limites presque sûres.

$B(t)$ y soit discontinu. C'est pour cela que nous avons supposé la fonction $B(t)$ continue et croissante, et il est évident que dans ce cas il n'y a qu'à prendre cette fonction comme nouveau paramètre pour être ramené au cas où $B(t) = t$.

Il faut alors prendre garde que ce changement de paramètre modifie la loi de probabilité dont dépend le choix des t_n ; c'est avec la loi de probabilité ainsi transformée que l'on pourra considérer la nouvelle fonction $X(t)$ comme un modèle de mouvement brownien linéaire pour lequel on ait $B(t) = t$.

Il peut être utile de préciser la loi dont dépend le choix des t_n de manière que la définition de ce que nous appelons un modèle ne dépende d'aucun élément arbitraire. Le plus simple est de supposer que chaque t_n soit choisi indépendamment des autres, et avec une probabilité uniformément répartie de zéro à un. On peut montrer que : *la notion de modèle de mouvement brownien linéaire ainsi obtenue n'est pas changée si l'on remplace cette loi de répartition uniforme par une autre loi absolument continue pour laquelle la densité de probabilité soit comprise entre deux nombres positifs.*

Nous n'indiquerons que le principe de la démonstration. Même si l'on suppose seulement que $B_n(1)$ tend presque sûrement vers $B(1)$, il en résulte que, pour tout t compris entre 0 et 1, $B_n(t)$ a presque sûrement une limite $B(t)$; autrement les oscillations de $B_n(t)$, qui dépendraient du choix des points de division entre 0 et t plus que de leur fréquence, ne seraient pas presque sûrement compensées par celles de $B_n(1) - B_n(t)$, qui dépendent des points de division choisis entre t et 1.

Il en résulte évidemment que l'on peut augmenter dans un rapport déterminé la probabilité d'un des intervalles $(0, t)$ et $(t, 1)$ et diminuer en conséquence celle de l'autre; cela ne peut pas empêcher que $B_n(t)$ et $B_n(1) - B_n(t)$ tendent respectivement, et presque sûrement, vers $B(t)$ et $B(1) - B(t)$; donc $B_n(1)$ vers $B(1)$.

On peut raisonner de la même manière pour n'importe quelle division de l'intervalle $(0, 1)$ en intervalles partiels, et un passage à la limite facile conduit au résultat énoncé.

Par suite, même si l'on précise la définition de modèle de mouvement de brownien linéaire par la condition que pour le choix de chaque t_n la probabilité soit répartie d'une manière uniforme, si l'on trouve pour $B_n(t)$ une limite presque sûre $T = B(t)$, pourvu que tous les rapports $\Delta T / \Delta t$ soient compris entre deux nombres positifs, le changement de variable qui consiste à prendre T comme nouvelle variable est légitime. On est ainsi ramené au cas où $B(t) = t$.

5°. La convergence de $B_n = B_n(1)$ vers $B = B(1)$ est bien entendu presque sûre, mais non sûre. Désignons par β_n la borne inférieure de B_n , et

par $\bar{\beta}_n$ sa borne supérieure, quand on fait varier les points de division. On a, au sujet de ces nombres, les résultats suivants :

THÉORÈME 8. *Pour n'importe quel modèle de mouvement brownien, β_n tend vers zéro, pour n infini.*

THÉORÈME 9. *Pour la fonction aléatoire $X(t)$ du schéma du mouvement brownien linéaire, il est presque sûr que $\bar{\beta}_n$ augmente indéfiniment avec n .*

Le premier de ces théorèmes résulte de ce qu'un modèle de mouvement brownien linéaire est nécessairement une fonction continue $X(t)$. On peut alors, si n est assez grand, définir entre $X(0)$ et $X(1)$ une suite de nombres croissants x_1, x_2, \dots, x_{n-1} telle que la somme des carrés des intervalles ainsi séparés soit arbitrairement petite, puis définir entre zéro et un des nombres croissants t_1, t_2, \dots, t_{n-1} tels que $X(t_v) = x_v$ ($v = 1, 2, \dots, n-1$). On obtient ainsi pour B_n une valeur arbitrairement petite, c. q. f. d.

On voit même aisément qu'on peut prendre pour les t_n n'importe quel ensemble dénombrable et partout dense entre zéro et un, donné d'avance; il suffit de les ranger dans un ordre convenable pour que B_n tende vers zéro (ou vers n'importe quelle valeur donnée entre zéro et B).

Pour démontrer le théorème 9, observons que, si les points de division sont assez nombreux, les valeurs des accroissements ΔX se répartissant suivant leur probabilité théorique, on aura avec une probabilité supérieure à $1 - \epsilon/2$ des intervalles de longueur totale supérieure à $k > 0$ pour lesquelles $(\Delta X)^2 > 2c\Delta t$ (c étant arbitrairement grand, ϵ arbitrairement petit, et k déterminé en fonction de c). Conservant ces intervalles, et subdivisant les autres, on arrivera de nouveau à trouver une fraction supérieure à k de la longueur de chacun d'eux pour laquelle on aura, pour les nouveaux intervalles obtenus, $(\Delta X)^2 > 2c\Delta t$, et cela en exceptant des cas de probabilité totale inférieure à $\epsilon/4$. Prenons alors pour p un entier tel que $(1 - k)^p < \frac{1}{2}$. Après p opérations analogues, sauf dans des cas de probabilité inférieure à

$$\epsilon/2 + \epsilon/4 + \dots + \epsilon/2^p < \epsilon,$$

on aura obtenu une division de l'intervalle $(0, 1)$ en intervalles partiels pour laquelle plus de la moitié de la longueur totale sera constituée par des intervalles partiels tels que $(\Delta X)^2 > 2c\Delta t$; donc $\Sigma(\Delta X)^2 > c$, ce qui démontre le théorème 9.

Le résultat ainsi obtenu pour la fonction aléatoire $X(t)$ n'est pas, comme dans le cas du théorème 8, applicable à tous les modèles de mouvement brownien linéaire. On peut définir de tels modèles pour lesquels on a toujours $(\Delta X)^2 \leq c\Delta t$, donc $B_n \leq c$, c étant une constante suffisamment grande.

Prenons maintenant pour n une fonction lentement croissante $n(h)$ d'un entier h (par exemple la partie entière de $\log \log \log h$), et supposons qu'à chaque valeur de h on fasse correspondre $n - 1$ points de division choisis au hasard, d'où résultera une valeur de $B_{n(h)} = B'_h$. Le grand nombre d'expériences ainsi faites pour une même valeur de n conduira à trouver, pour B_n , des nombres remplissant l'intervalle $(\beta_n, \bar{\beta}_n)$, et, si $n(h)$ croît assez lentement, la suite des B'_h aura presque sûrement pour valeurs limites tous les nombres de l'intervalle $(0, \bar{\beta})$, $\bar{\beta}$ étant la limite de $\bar{\beta}_n$.

D'après cette remarque, même s'il est possible de généraliser le résultat obtenu au sujet de la convergence presque sûre de B_n vers B , on ne peut pas l'appliquer sans aucune restriction relative au choix des modes de division de l'intervalle $(0, 1)$ successivement considérés. Nous ne savons pas, notamment, s'il serait suffisant que le nombre des points de division soit constamment croissant pour que la convergence de B_n vers B soit presque sûre.

6°. L'existence de fonctions qui soient des modèles de mouvement brownien linéaire n'était pas évidente a priori. Au point de vue idéaliste, elle résulte du théorème 7. Mais ce théorème ne nous donne aucun moyen de nommer une telle fonction; c'est ce que nous allons faire maintenant.

Pour cela nous nous inspirerons de ce que fait le hasard; nous chercherons à l'imiter. M. Borel a montré, qu'on ne peut pas, d'une manière générale, imiter le hasard; si l'on imite certains caractères d'une suite de nombres choisis au hasard, on en omet nécessairement d'autres. Mais si l'on porte son attention sur certaines conditions bien déterminées (ici celles qui interviennent dans la démonstration du théorème 7), on peut, à ce point de vue spécial, imiter le hasard.

Nous prendrons d'abord, comme modèle d'une suite de nombres $x_1, x_2, \dots, x_n, \dots$, choisis au hasard entre zéro et un, la suite des parties fractionnaires des nombres $n/\log n$. Elle présente ce caractère que, pourvu que $(n' - n)/\log n$ augmente indéfiniment avec n , les x_v d'indices compris entre n et n' se répartissent uniformément entre zéro et un, la fréquence de ceux qui sont compris entre zéro et x tendant nécessairement vers x . On imite le hasard en ce qui concerne l'uniformité de sa répartition; mais on ne l'imites pas dans ses caprices; une suite de nombres effectivement choisis au hasard ne serait pas constituée par des suites partielles de nombres croissant régulièrement de 0 à 1.

L'uniformité de la répartition serait encore mieux réalisée si l'on prenait n au lieu de $n/\log n$ (il suffirait que $n' - n$ augmente indéfiniment pour obtenir une répartition uniforme des termes d'indices compris entre n et n'). Mais il y aurait entre x_n et x_{2n} (qui serait la partie fractionnaire de $2x_n$) une

corrélation que nous évitons en prenant pour x_n la partie fractionnaire de $n/\log n$; il n'y a alors aucune corrélation entre x_n et x_{2n} , ni, plus généralement, entre x_n et $x_{n'}$, pour $n' = 2^b n$.

Posons alors $x_n = F(\xi_n)$, $F(\xi)$ désignant la fonction de répartition de la loi de Gauss; la suite des ξ_n sera un modèle de suite de variables gaussiennes choisies au hasard. Or, d'après le § 1, la détermination successive des nombres

$$X(1), X(\tfrac{1}{2}), X(\tfrac{1}{4}), X(\tfrac{3}{4}), X(\tfrac{1}{8}), X(\tfrac{3}{8}), \dots,$$

qui aboutit à la détermination de $X(t)$, dépend de variables gaussiennes choisies successivement. Il suffit de choisir la suite des ξ_n qui vient d'être définie pour obtenir un modèle de mouvement brownien linéaire; nous nous contenterons d'indiquer ce résultat sans démonstration.

7°. L'extension des résultats précédents au cas du mouvement brownien plan est immédiate. Nous désignerons ici par Δl la longueur de la corde $A(t)A(t + \Delta t)$, et ferons correspondre à chaque ligne polygonale à n côtés inscrite dans l'arc $A(0)A(1)$ la somme

$$(34) \quad B_n = \Sigma(\Delta l)^2 = \Sigma[(\Delta X)^2 + (\Delta Y)^2].$$

Les résultats obtenus pour le mouvement brownien linéaire s'appliquant séparément à $X(t)$ et $Y(t)$, on voit qu'ici B_n tend vers 2 [ou vers $2t$ si l'on considère l'arc $A(0)A(t)$] sous les conditions qui, dans le cas du mouvement linéaire, assurent la convergence vers 1 (ou vers t). La limite obtenue pourra toujours être appelée *mesure de l'oscillation brownienne* (plane).

On peut aussi se proposer de définir des *modèles de mouvement brownien plan*. Il faut d'abord supposer qu'il y ait, pour chacune des sommes $\Sigma(\Delta X)^2$ et $\Sigma(\Delta Y)^2$ relatives à chaque intervalle $(0, t)$, convergence presque sûre vers t . Mais cette condition est trop peu restrictive; elle pourrait être réalisée en prenant $Y(t) = X(t)$. Le mouvement serait rectiligne, et donnerait une bien mauvaise idée du mouvement brownien plan. Il est alors indiqué d'ajouter une condition d'indépendance de $X(t)$ et $Y(t)$; ce sera que $\Sigma \Delta X \Delta Y$ tende vers zéro, dans les conditions indiquées à propos de la convergence de $\Sigma(\Delta X)^2$ vers l'unité [ou vers t , s'il s'agit de l'arc $A(0)A(t)$]. Cette condition, équivalant à celle que la mesure de l'agitation brownienne linéaire soit toujours bien définie et ait la même valeur en projection sur n'importe quelle droite du plan, est évidemment réalisée, avec une probabilité unité, dans le schéma stochastique du mouvement brownien plan.

L'extension du théorème 9 au mouvement brownien plan est évidente. Il n'en est pas de même du théorème 8. La somme (34) n'est en effet très petite que si les deux termes $\Sigma(\Delta X)^2$ et $\Sigma(\Delta Y)^2$ sont tous les deux très petits. Or,

si chacun d'eux peut être rendu très petit, il n'est pas évident qu'ils peuvent être rendus simultanément très petits. Nous ne ferons que signaler ici cette question.

5. L'aire limitée par la courbe C . 1°. Etudiant maintenant le mouvement plan, nous désignerons par $S(t)$ l'aire comprise entre l'arc $A(0)A(t)$ et sa corde, les conventions de signes étant celles que l'on fait pour représenter une aire par une intégrale curviligne étendue à son contour; nous écrirons S au lieu de $S(1)$.

L'intégrale qui représente l'aire ne sera pas une intégrale au sens de Riemann, mais une intégrale stochastique, analogue à certains points de vue à l'intégrale B du § 4. Elle pourra être définie, non comme limite sûre d'une somme, mais comme limite en probabilité, ou en moyenne quadratique, ou encore comme limite presque sûre.

Comme pour l'étude de B_n , nous allons commencer par un cas simple en considérant les lignes polygonales L'_n , inscrites dans l'arc $A(0)A(1)$, ayant chacune pour sommets les points de cotes multiples 2^{-n} . Désignons par S'_n l'aire comprise entre L'_n et L'_{n+1} . Elle est la somme de 2^n triangles, dont les aires sont des variables aléatoires indépendantes les unes; nous préciserons plus loin la loi dont elles dépendent; il suffit d'observer ici qu'elles ont une valeur probable nulle, celle de leurs carrés étant $1/2^{2n+3}$. On en déduit

$$\mathcal{E}\{S'_n\} = 0, \quad \mathcal{E}\{S'^2_n\} = 1/2^{2n+3}.$$

La série Σ'_n , qui représente S , est donc convergente en moyenne quadratique. Quoique ces termes ne soient pas indépendants, il est facile d'établir sa convergence presque sûre. On peut, par exemple, utiliser l'inégalité de Tchebycheff, qui donne

$$Pr\{|S'_n| > 1/2^{(n+6)/4}\} < 1/2^{n/2}.$$

Cette probabilité étant le terme d'une série convergente, il existe presque sûrement un nombre N tel que, pour $n > N$, on ait

$$S'_n \leq 1/2^{(n+6)/4},$$

ce qui établit le résultat annoncé.

2°. Considérons maintenant, comme au 2° du § 4, une suite de nombres $t_1, t_2, \dots, t_n, \dots$, compris entre zéro et un et formant un ensemble partout dense dans cet intervalle. Nous désignerons par L_n la ligne brisée allant de $A(0)$ à $A(1)$ et ayant comme sommets intermédiaires les points A_1, A_2, \dots, A_{n-1} [en écrivant A_h au lieu de $A(t_h)$], rangés dans l'ordre des t croissants. Nous désignerons par S_n l'aire comprise entre L_n et la corde

$A(0)A(1)$, et par T_n la différence $S_{n+1} - S_n$, qui est l'aire d'un triangle ayant pour sommet A_n et pour base un côté $A_n'A_n''$ de L_n . Nous désignerons sa longueur par l_n , par t_n' et t_n'' les cotes de A_n' et A_n'' , et poserons

$$\tau_n = t_n'' - t_n', \quad \tau_n' = t_n - t_n', \quad \tau_n'' = t_n'' - t_n.$$

Si \mathcal{E}_{n-1} désigne une valeur probable calculée en connaissant A_1, A_2, \dots, A_{n-1} , on a

$$(35) \quad \begin{cases} \mathcal{E}_{n-1}\{T_n\} = 0, \quad \mathcal{E}_{n-1}\{T_n^2\} = l_n^2 \frac{\tau_n' \tau_n''}{4\tau_n}, \\ \mathcal{E}\{T_n^2\} = \frac{\tau_n' \tau_n''}{4\tau_n} \mathcal{E}\{l_n^2\} = \frac{\tau_n' \tau_n''}{2} = \frac{1}{4}(b_n - b_{n+1}), \end{cases}$$

b_n ayant la même signification qu'au § 4, de sorte que la formule (28) est toujours applicable. D'après la première de ces formules, on peut appliquer l'inégalité de A. Kolmogoroff à la somme ΣT_v . Il vient ainsi

$$Pr\left\{ \max_{0 < h \leq p} |S_{n+h} - S_n| \geq \frac{c}{2} \sqrt{b_n - b_{n+p}} \right\} \leq \frac{1}{c^2},$$

et, en faisant augmenter p indéfiniment

$$(36) \quad Pr\left\{ \max_{v > n} |S_v - S_n| \geq \epsilon' \right\} \leq \epsilon \quad \left(\epsilon' = \frac{c}{2} \sqrt{b_n}, \quad \epsilon = \frac{1}{c^2} \right).$$

Comme b_n tend vers zéro pour n infini, ϵ et ϵ' peuvent être rendus simultanément arbitrairement petits. La convergence presque sûre de la suite des S_n en résulte.¹¹

3°. Montrons maintenant que, si l'on remplace la suite des t_n par une autre suite analogue $\bar{t}_1, \bar{t}_2, \dots, \bar{t}_n, \dots$, les deux expressions S et \bar{S} successivement obtenues pour l'aire étudiée sont presque sûrement les mêmes. Il suffit évidemment de montrer que les aires polygonales S_n et \bar{S}_n dont elles sont les limites sont infiniment peu différentes en probabilité, et pour cela de montrer qu'elles sont l'une et l'autre infiniment peu différentes en probabilité de l'aire S_n'' limitée par la ligne polygonale inscrite dans C ayant pour sommets tous les points de cotes $t_1, t_2, \dots, t_{n-1}, \bar{t}_1, \bar{t}_2, \bar{t}_{n-1}$. Pour S_n , par exemple, cela résulte de la formule (36), qui s'applique évidemment à tout mode de subdivision de l'arc $A(0)A(1)$ commençant par les points de cotes t_1, t_2, \dots, t_{n-1} .

On peut aussi, en utilisant les formules (35), montrer que les moyennes quadratiques de $S_n - S_n''$ et $\bar{S}_n - S_n''$ sont infiniment petites.

¹¹ Au point de vue de l'uniformité de la convergence par rapport au choix de la suite des t_n , il faut noter que ϵ et ϵ' ne dépendent que de b_n , lui-même borné supérieurement par ϵ_n [formule (27)].

4°. Introduisons maintenant le hasard dans le choix des t_n . Les raisonnements étant identiques à ceux faits à propos des B_n , nous ne ferons qu'en rappeler les grandes lignes. C'est pour n'avoir pas à les recommencer que nous avons introduit à propos de l'étude de B_n l'expression générale (33), dont B_n n'était qu'une forme particulière; il faut seulement noter que S_n dépend des deux fonctions aléatoires $X(t)$ et $Y(t)$; mais cela ne change rien au raisonnement fait à propos de l'expression (33), et le résultat obtenu à cet endroit s'applique à S_n . Nous n'avons donc pas à craindre que nos raisonnements introduisent des ensembles non mesurables ou des probabilités non déterminées. La probabilité de la convergence de S_n vers S est bien déterminée, et il importe peu, pour la calculer, qu'on fasse d'abord les expériences qui déterminent les t_n , puis celles qui déterminent C , ou l'inverse. Or nous savons que, pourvu qu'il soit presque sûr que la suite des t_n est partout dense, il est presque sûr que S_n tend vers S . En disant que, pour une courbe C déterminée, l'aire S est *stochastiquement définie* si, les points t_n étant choisis au hasard, S_n tend presque sûrement vers une limite non aléatoire S , nous voyons que :

THÉOREME 10. *Le schéma aléatoire du mouvement brownien plan conduit, avec une probabilité unité, à une courbe C pour laquelle l'aire S est stochastiquement bien définie.*¹²

5°. La question se pose naturellement de déterminer la loi dont dépend S . Nous traiterons d'abord un problème plus élémentaire : *déterminer la loi dont dépend l'aire d'un triangle inscrit dans C , ses sommets ayant des cotes données.*

Nous désignerons ces cotes par $t - \tau'$, t , $t + \tau''$. L'aire est évidemment de la forme $\frac{1}{2} \sqrt{\tau' \tau''} T$, la nature de la variable aléatoire T étant indépendante de t , τ' et τ'' . Si λ désigne la longueur d'un vecteur gaussien réduit, et si η est une variable gaussienne réduite, la longueur du côté $A(t - \tau')A(t)$ du triangle étudié est de la forme $\lambda \sqrt{\tau'}$, et celle de la hauteur perpendiculaire à ce côté est $\eta \sqrt{\tau''}$; λ et η sont indépendants, et $T = \lambda \eta$.

Calculons les moments de la variable aléatoire T . On a évidemment

$$(37) \quad \begin{cases} E_{2p+1} = \mathcal{E}\{T^{2p+1}\} = \mathcal{E}\{\lambda^{2p+1}\} \mathcal{E}\{\eta^{2p+1}\} = 0, \\ E_{2p} = \mathcal{E}\{\lambda^{2p}\} \mathcal{E}\{\eta^{2p}\} = 2^p \Gamma(p+1) \cdot 1 \cdot 3 \cdot 5 \cdots (2p-1) = (2p)!, \end{cases}$$

¹² Bien entendu, l'aire S n'est définie que stochastiquement. On peut développer à ce sujet des remarques analogues à celles qui nous ont conduit au théorème 9. Le résultat est que C a presque sûrement la propriété suivante: on peut définir la suite des t_n de manière que S_n ne tende pas vers S , et même de manière que S_n ait n'importe quelle limite donnée, finie ou infinie.

et par suite

$$(38) \quad \phi(z) = \mathcal{E}\{e^{izT}\} = \sum_0^{\infty} (-z^2)^p = \frac{1}{1+z^2} \quad (|z| < 1).$$

Le fait qu'on trouve pour $\phi(z)$ une série entière à rayon de convergence positif suffit, comme on sait, pour être assuré que la fonction caractéristique de la loi étudiée est bien celle définie sur tout l'axe réel par le prolongement analytique de cette fonction. Il s'agit donc de la première loi de Laplace, c'est-à-dire de la loi symétrique définie par

$$(39) \quad \text{Pr}\{|T| > x\} = e^{-x}, \quad (x > 0).$$

6°. Etudions maintenant la loi dont dépend $S(1)$; à cause de la similitude stochastique entre la courbe C et ses parties, $S(t)/t$ dépend de la même loi. Cette variable aléatoire étant en corrélation avec la longueur $L(t) = \lambda\sqrt{t}$ de la corde $A(0)A(t)$, nous étudierons la fonction de répartition à deux variables, évidemment indépendante de t

$$F(\alpha, \rho) = \text{Pr}\{S(t) < \alpha t, L(t) < \rho\sqrt{t}\}.$$

Nous commencerons par admettre que les trois premières dérivées de cette fonction sont définies et continues, sauf peut-être pour $\alpha = 0$; cette hypothèse sera justifiée plus loin. Il résulte d'autre part de la manière dont S a été défini comme somme d'une série qui converge en moyenne quadratique (§ 5, 1°), que ses deux premiers moments sont finis.

Nous désignerons par $\xi\sqrt{dt}$ et $\eta\sqrt{dt}$ les deux composantes de $A(t)A(t+dt)$ suivant la direction $A(0)A(t)$ et la direction perpendiculaire; ξ et η sont deux variables gaussiennes réduites, indépendantes l'une de l'autre, et indépendantes de l'arc $A(0)A(t)$, et par suite de $S(t)$ et $L(t)$. De

$$(L + \delta L)^2 = (L + \xi\sqrt{dt})^2 + \eta^2 dt$$

[en écrivant L au lieu de $L(t)$], on déduit

$$(40) \quad \delta L = \xi\sqrt{dt} + \frac{\eta^2}{2L} dt + o(dt),$$

tandis que la variation de $S = S(t)$ est évidemment

$$(41) \quad \delta S = \frac{L\eta}{2} \sqrt{dt} + S_1(dt),$$

$S_1(dt)$ dépendant de la même loi que $S(dt)$, donc que $dtS(1)$; cette aire a sa valeur probable nulle, et est stochastiquement indépendante de l'arc $A(0)A(t)$, donc de $S(t)$ et $L(t)$, mais non de ξ et η .

Pour former une équation vérifiée par la fonction $F(\alpha, \rho)$, nous allons calculer de deux manières différentes la probabilité

$$(42) \quad P = \Pr\{S + \delta S < \alpha t, L + \delta L < \rho \sqrt{t}\}.$$

Une première évaluation de P repose sur la remarque que, $F(\alpha, \rho)$ ne dépendant pas du temps, on a

$$\Pr\{S + \delta S < \alpha_1(t + dt), L + \delta L < \rho_1 \sqrt{t + dt}\} = F(\alpha_1, \rho_1).$$

En définissant alors α_1 et ρ_1 par les formules

$$\alpha_1(t + dt) = \alpha t, \quad \rho_1 \sqrt{t + dt} = \rho \sqrt{t},$$

d'où l'on tire

$$\alpha_1 - \alpha \sim -\alpha \frac{dt}{t}, \quad \rho_1 - \rho \sim -\rho \frac{dt}{2t} \quad (dt \rightarrow 0),$$

il vient

$$(43) \quad P = F(\alpha_1, \rho_1) = F(\alpha, \rho) - \alpha \frac{dt}{t} F'_\alpha - \rho \frac{dt}{2t} F'_\rho + o(dt).$$

La seconde manière d'évaluer P repose sur les expressions (40) et (41) de δL et δS , qui montrent que l'expression

$$P_1 = \Pr \left\{ S < \left(\alpha - \frac{\eta L \sqrt{dt}}{2t} - \frac{S_1(dt)}{t} \right) t, L < \left(\rho - \xi \sqrt{\frac{dt}{t}} - \frac{\eta^2 dt}{2\rho t} \right) \sqrt{t} \right\}$$

est de la forme $P + o(dt)$. Pour évaluer P_1 , nous poserons

$$\alpha - \frac{S_1(dt)}{t} = \alpha', \quad \rho - \xi \sqrt{\frac{dt}{t}} - \frac{\eta^2 dt}{2\rho t} = \rho',$$

et désignerons par P'_1 , au lieu de P_1 , une probabilité conditionnelle évaluée en supposant connus ξ , η , et $S_1(dt)$; P_1 est sa valeur probable. On a évidemment

$$\begin{aligned} P'_1 &= \int_0^{\rho'} F'_\rho(\alpha' - \frac{\eta\lambda}{2} \sqrt{\frac{dt}{t}}, \lambda) d\lambda \\ &= \int_0^{\rho'} \left[F'_\rho(\alpha', \lambda) - \frac{\eta\lambda}{2} \sqrt{\frac{dt}{t}} F''_{\alpha\rho} + \frac{\eta^2 \lambda^2}{8} \frac{dt}{t} F'''_{\alpha^2\rho} + K \eta^3 \left(\frac{dt}{t} \right)^{3/2} \right] d\lambda, \end{aligned}$$

$|K|$ ayant une borne supérieure, fonction de α' et ρ' seulement. Comme

$$\mathcal{E}\{\eta\} = 0, \quad \mathcal{E}\{\eta^2\} = 1, \quad \mathcal{E}\{|\eta^3|\} < \infty,$$

il vient

$$P_1 = \mathcal{E}\{P'_1\} = \mathcal{E}\{F(\alpha', \rho')\} + \frac{dt}{8t} \int_0^{\rho'} \lambda^2 F'''_{\alpha^2\rho}(\alpha, \lambda) d\lambda + o(dt).$$

Tenant compte d'autre part de

$$F(\alpha', \rho') = F(\alpha, \rho) - \frac{S_1(dt)}{t} F'_\alpha - \left(\xi \sqrt{\frac{dt}{t}} + \frac{\eta^2 dt}{2\rho t} \right) F'_\rho + \xi^2 \frac{dt}{2t} F''_{\rho^2} + K_1 o(dt),$$

où $|K_1|$ est borné par un polynôme en $|\xi|$, $|\eta|$ et $|S_1(dt)/t|$ dont la valeur probable est finie, et où l'expression désignée par $o(dt)$ n'est pas aléatoire, il vient

$$P_1 = F - \frac{dt}{2\rho t} F'_\rho + \frac{dt}{2t} F''_{\rho^2} + \frac{dt}{8t} \int_0^\rho \lambda^2 F'''_{\alpha^2 \rho}(\alpha, \lambda) d\lambda + o(dt).$$

Comme $P_1 = P + o(dt)$, la comparaison de cette formule et de la formule (43) donne

$$(44) \quad 2\alpha F'_\alpha + \left(\rho - \frac{1}{\rho} \right) F'_\rho + F''_{\rho^2} + \int_0^\rho \frac{\lambda^2}{4} F'''_{\alpha^2 \rho}(\alpha, \lambda) d\lambda = 0,$$

d'où, en dérivant par rapport à ρ et posant $F'_\rho = G$,

$$(45) \quad \left(1 + \frac{1}{\rho^2} \right) G + 2\alpha G'_\alpha + \left(\rho - \frac{1}{\rho} \right) G'_\rho + G''_{\rho^2} + \frac{\rho^2}{4} G''_{\alpha^2} = 0.$$

7°. THÉORÈME 11. $F(\alpha, \rho)$ est la seule solution de l'équation (44) qui soit une fonction de répartition.

L'équation (44) exprime en effet que $S(t)/t$ et $L(t)/\sqrt{t}$ dépendent d'une loi à deux variables indépendantes de t . Si, pour $t = t_0 > 0$, on prenait une loi initiale quelconque, la variation de S et celle de L étant ensuite définies par les formules (40) et (41), on aurait, au lieu de $F(\alpha, \rho)$, une fonction $F(\alpha, \rho, t)$ vérifiant une équation analogue à l'équation (44), mais où il y aurait un second membre $2tF'_t$. Une solution de cette équation étant bien déterminée par sa valeur initiale (pour $t = t_0$), l'équation (44) exprime bien la condition nécessaire et suffisante pour que la loi de probabilité considérée ne varie pas avec t .

Supposons alors qu'on ait deux solutions différentes F_1 et F_2 de l'équation (44), qui soient des fonctions de répartition. On peut leur faire correspondre deux systèmes de variables aléatoires S_1, L_1 et S_2, L_2 qui, pour deux courbes C_1 et C_2 dépendant de processus convenablement déterminés pour t variant de zéro à un, représentent l'aire comprise entre l'arc $A(0)A(1)$ et sa corde, et la longueur de cette corde; le système S_1, L_1 dépendra de la loi définie par F_1 ; S_2, L_2 de celle définie par F_2 .

Déplaçons maintenant la figure sur laquelle est tracée une de ces courbes, de manière que, pour les deux courbes considérées, $A(1)$ ait la même position, et $A(1)A(0)$ la même orientation; les deux origines A_0 et A_1 des deux courbes seront ainsi sur une même demi-droite issue de leur extrémité commune $A(1)$. Faisant ensuite varier t à partir de la valeur 1, on prolongera C_1 et C_2 par la même courbe C dépendant du schéma stochastique du mouvement brownien. Alors le système $S_1(t)/t$, $L_1(t)/\sqrt{t}$ ne cessera pas de dépendre de la loi définie par F_1 ; le système $S_2(t)/t$, $L_2(t)/\sqrt{t}$, dépendra de même de celle définie par F_2 .

Or $|L_1 - L_2|$ est borné supérieurement par la distance A_1A_2 , et $2(S_2 - S_1)/A_1A_2$ représente la distance du point $A(t)$ à A_1A_2 ; c'est une variable gaussienne de paramètre $\sqrt{t-1}$. Il en résulte qu'asymptotiquement, pour t infini, les deux systèmes de variables considérés sont confondus; en termes précis, les différences

$$\frac{S_1 - S_2}{t}, \quad \frac{L_1 - L_2}{\sqrt{t}},$$

tendent en probabilité vers zéro. Ils dépendent donc, à la limite, de la même loi de probabilité; donc $F_1 = F_2$, c. q. f. d.

8°. Quoique les équations (44) et (45) résolvent théoriquement le problème posé, nous n'avons pas pu obtenir l'expression explicite de $F(\alpha, \rho)$; peut-être n'existe-t-il aucune expression simple de cette fonction. Dans ces conditions il peut être utile d'indiquer d'autres méthodes qui permettent d'étudier la variable aléatoire S et sa corrélation avec L . Dans ce qui suit, L et S ne désigneront plus $L(t)$ et $S(t)$, mais $L(1)$ et $S(1)$.

On peut d'abord calculer les moments de la loi à deux variables L et S .¹³ Le moment

$$\mathcal{E}\{L^p S^q\} = E_{p,q}$$

étant évidemment nul si q est impair, il suffit de calculer ceux dont le second indice est pair. Le calcul repose sur la formule (41), formule exacte où l'on peut remplacer t et dt par un. En désignant par L et L_1 les longueurs des cordes $A(0)A(1)$ et $A(1)A(2)$, et par ϕ leur angle, qui est une variable choisie au hasard entre $-\pi$ et $+\pi$, cette formule prend la forme

$$(46) \quad S(2) = S + S_1 + \frac{1}{2} L L_1 \sin \phi.$$

¹³ Tous ces moments sont finis; cela résulte évidemment de

$$E_{p,q} \leq \frac{1}{2} (E_{2p,0} + E_{0,2q}),$$

et de ce que, comme nous le verrons plus loin, la probabilité des grandes valeurs de $|S|$ décroît comme une exponentielle.

Or ϕ d'une part, L et S d'autre part, L_1 et S_1 en dernier lieu, constituent des groupes de variables indépendants les uns des autres; la loi à deux variables L et S est la même que celle dont dépend L_1 , S_1 , et, par un changement d'unité, détermine celle dont dépend le système $S(2)$, $L(2)$. On a ainsi

$$\begin{aligned}\mathcal{E}\{S^2(2)\} &= 4E_{0,2} = \mathcal{E}\{S^2 + S_1^2 + \tfrac{1}{4}L^2L_1^2\sin^2\phi\} \\ &= 2E_{0,2} + \tfrac{1}{4}\mathcal{E}\{L^2\}\mathcal{E}\{L_1^2\}\mathcal{E}\{\sin^2\phi\} = 2E_{0,2} + \tfrac{1}{2},\end{aligned}$$

et par suite

$$(47) \quad E_{0,2} = \mathcal{E}\{S^2\} = \tfrac{1}{4}, \quad \mathcal{E}\{S^2(t)\} = \frac{t^2}{4}.$$

On détermine ensuite $E_{2,2}$ en calculant $\mathcal{E}\{L^2(2)S^2(2)\}$ à l'aide de la formule (46) et de

$$(46') \quad L^2(2) = L_1^2 + L_2^2 + 2L_1L_2\cos\phi,$$

puis $E_{0,4}$, et ainsi de suite. Tous les moments sont ainsi obtenus sans difficulté, mais successivement, le calcul dépendant chaque fois de moments antérieurement calculés. C'est donc une méthode de récurrence, et l'expression générale de ces moments peut être difficile à obtenir.

9°. Une autre méthode repose sur les remarques sur l'interpolation faites plus haut (§ 3, 1°). La détermination de l'arc $A(0)A(1)$ résulte de la détermination de vecteurs gaussiens indépendants, qui définissent successivement $A(1)$, puis $A(\frac{1}{2})$, $A(\frac{1}{4})$, $A(\frac{3}{4})$, et ainsi de suite. Désignons par C_0 , $X_0(t)$, $Y_0(t)$, S_0 , ce que deviennent respectivement la courbe C et les grandeurs $X(t)$, $Y(t)$ et S , quand on remplace par zéro la longueur du premier de ces vecteurs, sans changer les autres. Si l'on a orienté l'axe des x parallèlement à $A(0)A(1)$, on a évidemment

$$Y(t) = Y_0(t), \quad X(t) = X_0(t) + tL \quad (0 \leq t \leq 1),$$

et par suite

$$(48) \quad S = \int_0^1 Y(t) dX(t) = S_0 + I_1L \quad [I_1 = \int_0^1 Y(t) dt].$$

Comme S_0 et I_1 ne dépendent que de l'orientation du premier des vecteurs successivement choisis, et des vecteurs suivants, l'ensemble des variables S_0 et I_1 est indépendant de L . Dans ces conditions la formule (48) définit bien la nature de la corrélation entre L et S ; elle montre notamment que le moment conditionnel $\mathcal{E}'\{S^p\}$, calculé dans l'hypothèse $L = \lambda$, est un polynôme de degré p en λ (évidemment nul si p est impair, et pair si p est pair).

Pour préciser ces renseignements, on peut chercher à définir les lois dont dépendent I_1 et S_0 et la corrélation entre ces variables. $Y(t)$ étant une somme de termes gaussiens de la forme $\eta\sqrt{dt}$, I_1 est de la forme

$$I_1 = \int_0^1 \eta(1-t) \sqrt{dt}$$

les η étant liés par la relation

$$Y_1 \equiv Y(1) \equiv \int_0^1 \eta \sqrt{dt} = 0.$$

Sans cette relation, la loi à deux variables Y_1 et I_1 serait une loi de Gauss, bien déterminée par ses moments du second ordre

$$\mathcal{E}\{I_1^2\} = \int_0^1 (1-t)^2 dt = \frac{1}{3}, \quad \mathcal{E}\{I_1 Y_1\} = \int_0^1 (1-t) dt = \frac{1}{2}, \quad \mathcal{E}\{Y_1^2\} = 1.$$

Par suite, dans l'hypothèse $Y_1 = 0$, I_1 est de la forme $\eta_1/2\sqrt{3}$, η_1 étant une variable gaussienne réduite.¹⁴ Le produit $I_1 L$ est alors de la forme $T/2\sqrt{3}$, T dépendant de la loi définie par la formule (39).

Pour l'étude de la loi à deux variables I_1, S_0 , on peut former une équation aux dérivées partielles analogue à l'équation (45) relative aux variables L et S . On peut aussi calculer ses moments. Observons seulement que, quand $Y(t)$ et par suite I_1 sont connus, S_0 dépend d'une loi symétrique; on a donc, si p est impair

$$(49) \quad \mathcal{E}\{S_0^p I_1^q\} = 0.$$

On déduit ensuite de la formule (48)

$$\mathcal{E}\{S^2\} = \mathcal{E}\{S_0^2\} + \frac{1}{4} \mathcal{E}\{I_1^2\} \mathcal{E}\{L^2\},$$

et par suite

$$(49') \quad \mathcal{E}\{S_0^2\} = \frac{1}{4} - \frac{1}{4} \cdot \frac{1}{12} \cdot 2 = \frac{5}{24}$$

(On remarque que $\mathcal{E}\{I_1^2\}$, calculé dans l'hypothèse $Y_1 = 0$, a la valeur $1/12$, et non $1/3$).

10°. La dernière des méthodes que nous voulons indiquer repose sur la remarque que la loi étudiée peut être définie comme limite de celle dont dépend l'aire S_n comprise entre une ligne polygonale L_n inscrite dans l'arc $A(0)A(1)$, et la corde de cet arc. Cela est bien évident, puisque S_n tend en probabilité vers $S(1)$. Or S_n est une somme de n triangles $A(0)A(t)A(t+dt)$ (t et $t+dt$ désignant les cotes de deux sommets consécutifs de L_n); il suffit donc d'étudier la loi dont dépend cette somme.

¹⁴ La manière la plus simple de calculer le coefficient numérique $\frac{1}{2\sqrt{3}}$ est sans doute de remarquer que $\mathcal{E}\left\{\left[\int_0^1 [Y_0(t) + tY_1]dt\right]^2\right\}$ a la valeur $\frac{1}{3}$ obtenue pour $\mathcal{E}\{I_1^2\}$ quand $Y_1 = Y(1)$ n'est pas supposé connu.

Désignons toujours par $\xi\sqrt{dt}$ et $\eta\sqrt{dt}$ les composantes de $A(t)A(t+dt)$ suivant la direction $A(0)A(t)$ et la direction perpendiculaire, et considérons la loi conditionnelle dont dépend S_n , et à la limite S , lorsqu'on connaît les ξ et les $|\eta|$, et par suite toutes les valeurs successivement prises par $L(t)$. L'aire S_n se présente sous la forme d'une somme $\frac{1}{2}\sum \pm L(t)|\eta|\sqrt{dt}$, les signes seuls étant indéterminés, et indépendants les uns des autres. En négligeant des cas très peu probables, le plus grand de ces termes est très petit par rapport à la somme (la vérification ne présente aucune difficulté). Il en résulte que la loi conditionnelle limite obtenue pour S est la loi de Gauss, c'est-à-dire que S est de la forme $\sigma\xi$, σ désignant la valeur quadratique moyenne de S , pour la loi conditionnelle; ξ est une variable gaussienne réduite; elle est donc indépendante de σ , et l'on est ramené à étudier la loi dont dépend l'expression

$$\sigma^2 = \lim \sum \frac{L^2 \eta^2 dt}{4} = \frac{1}{4} \int_0^1 L^2(t) dt,$$

la convergence considérée étant une convergence en probabilité.

Or $4\sigma^2$ est la somme des deux intégrales

$$J = \int_0^1 X^2(t) dt, \quad J_1 = \int_0^1 Y^2(t) dt,$$

indépendantes l'une de l'autre, et dépendant d'une même loi; on est ramené à étudier cette loi.

On peut appliquer à l'étude de la loi à deux variables J et $X = X(1)$ des méthodes analogues à celles appliquées à l'étude de S et L . Indiquons seulement sans démonstration qu'en posant

$$\frac{\partial}{\partial \alpha} \text{Pr}\{X < \alpha, J < \beta\} = H(\alpha, \beta),$$

on obtient l'équation aux dérivées partielles

$$(50) \quad H + \alpha H'_\alpha + (4\beta - 2\alpha^2) H'_\beta + H''_{\alpha^2} = 0,$$

qui joue un rôle analogue à celui de l'équation (45), mais qui est du type parabolique. On peut montrer, ici encore, qu'elle détermine complètement la loi étudiée.

On peut aussi calculer successivement les différents moments de la loi à deux variables J et X , et aussi montrer que la corrélation entre J et X est du second degré, c'est-à-dire que

$$J = J_0 + 2I_0X + X^2,$$

I_0 et J_0 étant indépendants de X .

L'étude de J est ainsi analogue à celle de S , mais à certains points de vue plus simple; on remarque que J est une intégrale de type classique, bien qu'elle dépende d'une fonction aléatoire; de plus elle ne dépend que de la seule fonction $X(t)$. Malgré cela la fonction H ne nous a pas plus que G paru susceptible d'avoir une expression simple.

La formule obtenue

$$(51) \quad S = \sigma \xi \quad (4\sigma^2 = J + J_1, \sigma > 0)$$

n'en est pas moins susceptible de donner des renseignements utiles sur la nature de la variable aléatoire S . Elle montre notamment que la fonction de répartition de S est continue et indéfiniment dérivable, sauf peut-être pour la valeur zéro de la variable; la loi qu'elle définit est symétrique. Ce sont en effet des propriétés qui sont nécessairement vérifiées par un produit de deux facteurs indépendants, si elles sont vérifiées par un des facteurs (ici ξ).

11°. Nous allons indiquer une autre conséquence de la formule (51): *au point de vue de l'ordre de grandeur de la probabilité des grandes valeurs de $|S|$, la loi dont dépend S est comparable à la première loi de Laplace.*

Majorons d'abord la probabilité des grandes valeurs de σ . On a

$$\begin{aligned} Pr\{\sigma > s\} &\leq Pr\{\text{Max}_{t \leq 1}(J, J_1) > 2s^2\} \leq 2Pr\{J > 2s^2\} \\ &\leq 2Pr\{\text{Max}_{t \leq 1}|X(t)| > s\sqrt{2}\} \leq 4Pr\{\text{Max}_{t \leq 1}X(t) > s\sqrt{2}\}, \end{aligned}$$

c'est-à-dire

$$Pr\{\sigma > s\} \leq 8Pr\{X > s\sqrt{2}\}.$$

Par suite, si $c\sqrt{2} > 1$, on a, pour s assez grand

$$(52) \quad Pr\{\sigma > s\} \leq e^{-s^2/2c^2} = Pr\{c\lambda > s\},$$

λ désignant toujours la longueur d'un vecteur gaussien réduit.

Or une inégalité de cette forme subsiste si l'on multiplie à la fois σ et λ par la variable $|\xi|$, non négative, et indépendante d'elles. Elle exprime en effet qu'on peut établir entre σ et λ une corrélation telle que l'on ait toujours $\sigma < c\lambda$. On en déduit

$$(53) \quad Pr\{|S| > s\} \leq Pr\{c|T| > s\} \quad (c\sqrt{2} > 1),$$

$T = \lambda\xi$ dépendant de la première loi de Laplace, comme nous l'avons vu plus haut.

Pour obtenir au contraire une borne supérieure du premier membre, remarquons que dans la formule (48), lorsque L et $Y(t)$ (donc I_1) sont connus, $X_0(t)$, et par suite S_0 , dépendent de lois symétriques. Il y a donc

une chance sur deux pour que S_0 soit du signe de I_1 , et que par suite on ait $|S| \geq \frac{1}{2} L |I_1|$. On en déduit

$$Pr\{|S| > s\} \geq \frac{1}{2} Pr\{L |I_1| > s\} = \frac{1}{2} Pr\left\{\frac{T}{2\sqrt{3}} > s\right\},$$

et par suite, pour c assez petit ($2c\sqrt{3} < 1$) et s assez grand

$$(54) \quad Pr\{|S| > s\} \geq Pr\{c' |T| > s\} \quad (2c\sqrt{3} < 1).$$

Il est alors probable qu'on peut déterminer une constante absolue c' (comprise entre $1/\sqrt{2}$ et $1/2\sqrt{3}$ ou égale à un de ces nombres) telle que la formule (53) s'applique pour $c > c'$ et la formule (54) pour $c < c'$.

12°. Nous allons maintenant établir le résultat annoncé plus haut concernant l'existence et la continuité des dérivées de $F(\alpha, \rho)$. Nous n'avons pu y arriver que par l'utilisation simultanée des différentes méthodes d'étude de cette fonction exposée ci dessus; mais nous ne serions pas surpris qu'il existe une démonstration plus simple.

Utilisons d'abord la formule (48) où, quand $Y(t)$ et par suite I_1 sont connus, S_0 dépend évidemment d'une loi continue [et même à fonction de répartition indéfiniment dérivable; on le voit aisément en étudiant l'influence d'un des paramètres indépendants qui définissent $X_0(t)$, par exemple $X_0(\frac{1}{2})$]. Il en résulte qu'il ne peut pas y avoir de relation linéaire entre S_0 et I_1 qui ait une probabilité positive d'être réalisée. La probabilité conditionnelle de $S < \alpha$, quand L a une valeur connue ρ , est donc une fonction continue de α et ρ ; comme on obtient évidemment $F(\alpha, \rho)$ en multipliant cette probabilité conditionnelle par $\rho e^{-\rho^2/2} d\rho$, qui est la probabilité de l'intervalle $d\rho$, et en intégrant par rapport à ρ , il en résulte que la dérivée $G = F'_\rho$ existe, et est une fonction continue de ρ ; au facteur $\rho e^{-\rho^2/2}$ près, elle représente la probabilité conditionnelle de $S < \alpha$, dans l'hypothèse $L = \rho$.

Utilisons maintenant la formule (51). L'hypothèse $L = \rho$ ne modifie pas le fait que, dans cette formule, ξ soit une variable gaussienne indépendante de σ . La probabilité conditionnelle de $S < \alpha$, évaluée dans cette hypothèse, est donc aussi, comme la probabilité non conditionnelle de la même inégalité, une fonction continue et indéfiniment dérivable de α , sauf peut-être pour $\alpha = 0$; G , et par suite F , sont donc aussi indéfiniment dérivables par rapport à α .

La formule (48), compte tenu de la remarque faite tout à l'heure sur la loi dont dépend S_0 quand I_1 est connu, donne aisément une autre démonstration de ce résultat.

Reportons nous maintenant au raisonnement par lequel nous avons établi

l'équation (44) ; nous allons montrer que la continuité de F et G et de toutes leurs dérivées par rapport à α sont des hypothèses suffisantes pour que ce raisonnement subsiste, avec quelques modifications. La dérivée F''_{ρ^2} a d'abord été introduite par le développement de l'expression

$$F\left(\alpha, \rho - \xi \sqrt{\frac{dt}{t}}\right) = F(\alpha, \rho) - \xi \sqrt{\frac{dt}{t}} F'_{\rho} + \frac{\xi^2 dt}{2t} F''_{\rho^2} + o(dt).$$

Si l'on n'est pas assuré a priori de son existence, il résulte seulement de la comparaison des deux expressions obtenues pour P et $P_1 = P + o(dt)$ que

$$\varepsilon \left\{ F\left(\alpha, \rho - \xi \sqrt{\frac{dt}{t}}\right) - F(\alpha, \rho) + \xi \sqrt{\frac{dt}{t}} F'_{\rho} \right\}$$

est de la forme $k dt/2t + o(dt)$, $k = k(\alpha, \rho)$, qui ne dépend que des valeurs de $F(\alpha, \lambda)$ pour λ très voisin de ρ , étant ce qu'on peut appeler une *dérivée seconde généralisée*. On peut alors écrire l'équation (44), à condition de remplacer F''_{ρ^2} par k .

D'après cette équation, k est une fonction continue de α et ρ . La différence

$$K(\alpha, \rho) = F(\alpha, \rho) - \int_0^{\rho} k(\alpha, \lambda) (\rho - \lambda) d\lambda$$

a une dérivée seconde généralisée nulle. Il en résulte qu'elle est linéaire en ρ . Si en effet il n'en était pas ainsi, on pourrait trouver une valeur ρ_0 de ρ telle que la courbe représentative de la fonction K (pour α constant) soit intérieure à une parabole qui la touche au point d'abscisse ρ_0 . L'expression

$$K(\alpha, \rho) - K(\alpha, \rho_0) - (\rho - \rho_0) K'_{\rho}(\alpha, \rho_0)$$

serait donc de signe constant et supérieure en valeur absolue à $c(\rho - \rho_0)^2/2$, c étant positif; la dérivée seconde généralisée apparaîtrait donc, d'après sa définition, comme une valeur moyenne d'une quantité de signe constant et supérieure à c en valeur absolue, et ne pourrait pas être nulle, comme nous l'avions supposé. La fonction $K(\alpha, \rho)$ est donc linéaire en ρ , et F admet une dérivée seconde F''_{ρ^2} , égale à k .

Il s'agit d'ailleurs d'un résultat général: si une fonction $f(x)$ admet une dérivée seconde généralisée, définie pour chaque point x par une formule du type

$$f''(x) = \lim \mathfrak{M} \left\{ \frac{2}{h^2} [f(x+h) - f(x) - hf'(x)] \right\},$$

\mathfrak{M} désignant une moyenne pondérée (par rapport à h), et les petites valeurs de h intervenant seules à la limite, et si $f''(x)$ continu, c' est une dérivée seconde au sens ordinaire.

Revenons à l'équation (44), dont l'exactitude est maintenant établie. Tous les termes autres que $F''\rho^2$ étant dérivables par rapport à ρ , ce terme l'est aussi. L'équation (45) en résulte, et, comme elle est du type elliptique, les fonctions $F(\alpha, \rho)$ et $G(\alpha, \rho)$ sont continues et indéfiniment dérivables, sauf peut-être pour $\alpha = 0$.

Le résultat subsiste d'ailleurs pour $\alpha = 0$. Si l'on se reporte à la formule (51), on peut remarquer que, si l'on remplaçait σ par la longueur d'un vecteur gaussien, le produit $\sigma\xi$ dépendrait de la première loi de Laplace, pour laquelle la dérivée seconde de la fonction de répartition est discontinue à l'origine. Mais il suffit, pour écarter la possibilité d'une telle discontinuité, de montrer que, dans le cas qui nous occupe, les petites valeurs de σ sont très peu probables. Cela résulte aisément de la définition de $4\sigma^2$, somme de termes tous positifs; la probabilité qu'ils soient tous très petits est excessivement petite (et cela aussi dans l'hypothèse où L est supposé connu).

Nous laisserons au lecteur de soin de préciser ce raisonnement, ce qui peut être fait de deux manières différentes. On peut montrer que, quel que soit c positif, on a

$$Pr\{\sigma < s\} = o(s^c) \quad (s \rightarrow 0),$$

et en déduire directement le résultat annoncé pour le produit $\sigma\xi$. On peut aussi établir seulement le résultat annoncé pour $c = 2$, et en déduire la continuité des dérivées qui figurent dans les équations (44) et (45). En raison du type elliptique de cette dernière équation, cela suffit pour conclure que $G(\alpha, \rho)$ est holomorphe pour toutes les valeurs réelles de α et toutes les valeurs positives de ρ .

6. La mesure superficielle de la courbe C . L'objet de ce paragraphe est de démontrer le théorème suivant.

THÉORÈME 12. *La courbe C est un ensemble de points dont la mesure superficielle est presque sûrement nulle.*

Le 1° de ce paragraphe est consacré à un résultat préliminaire; le 2° contient la démonstration du théorème 12. Le 3° et le 4° contiennent des remarques qui nous semblent de nature à faire comprendre, mieux peut-être que la démonstration, la véritable nature de ce théorème, et en tout cas préparent les généralisations qui seront l'objet du paragraphe suivant; le 3° contient en outre un théorème important par lui-même; il donne une condition nécessaire pour qu'une courbe remplisse une aire.

1°. Il suffit de considérer un arc $A(0)A(t)$ de la courbe C . Comme c'est un ensemble fermé, il a une mesure superficielle bien déterminée $\mu(t)$.

Nous nous proposons d'abord de démontrer que $\mu = \mu(1)$ est une variable aléatoire [donc aussi $\mu(t)$]. En d'autres termes, μ est une fonction mesurable de la variable U dont le choix au hasard entre zéro et un équivaut au choix de C , la notion de probabilité équivalant à la mesure sur l'axe des U .

La démonstration repose sur ce que μ est la limite en probabilité d'une suite de variables aléatoires; on sait qu'une telle limite est une variable aléatoire.

Désignons à cet effet par $\mu_n(\rho)$ la mesure de l'ensemble des points intérieurs à l'un au moins des $n + 1$ cercles de centres $A(0), A(1/n), A(2/n), \dots, A(1)$, et de même rayon ρ . C'est évidemment une variable aléatoire; nous allons montrer que, ρ tendant vers zéro, si n est une fonction de ρ de croissance assez rapide, $\mu_n(\rho)$ tend en probabilité vers μ ; ¹⁵ il en résultera que μ est une variable aléatoire.

D'une part $\mu_n(\rho)$ est borné supérieurement, d'une manière non aléatoire, par $\bar{\mu}(\rho)$, mesure du lieu des points M dont la distance $\delta(M)$ à l'arc $A(0)A(1)$ ne dépasse pas ρ . Or, en vertu d'une propriété connue des ensembles bornés et fermés, $\bar{\mu}(\rho)$ tend vers μ quand ρ tend vers zéro.

Désignons d'autre part par l_v la longueur du $v^{\text{ième}}$ côté de la ligne polygonale $A(0)A(1/n)A(2/n) \dots A(1)$, et par ρ_n le plus grand de ces côtés. On a évidemment

$$Pr\{\rho_n > \rho\} \leq \Sigma Pr\{l_v > \rho\} = ne^{-n\rho^2/2} = \epsilon(\rho)$$

et il suffit que n croisse assez rapidement quand ρ tend vers zéro pour que $\epsilon(\rho)$ tende vers zéro. Or, quand $\rho_n \leq \rho$, l'ensemble des $n + 1$ cercles considérés contient la courbe à son intérieur, et $\mu_n(\rho) \geq \mu$; il en est donc ainsi sauf dans des cas dont la probabilité est au plus égale à $\epsilon(\rho)$.

Finalement, $\mu_n(\rho)$ est au moins égal à μ , sauf dans des cas dont la probabilité tend vers zéro, et dans tous les cas au plus égal à $\bar{\mu}(\rho)$, qui tend vers μ ; la convergence en probabilité de $\mu_n(\rho)$ vers μ est ainsi établie.

On démontre de la même manière que, pour tout point M , $\delta(M)$, distance de ce point à la courbe, est une variable aléatoire; la probabilité que $\delta(M) = 0$, c'est-à-dire que M soit sur la courbe, est aussi bien déterminée, et est une fonction mesurable $\phi(M)$ du point M . Ces fonctions $\delta(M)$ et $\phi(M)$ sont en effet limites de celles obtenues en remplaçant la courbe par l'ensemble des cercles considérés dans le raisonnement précédent.

¹⁵ La condition que n croisse assez rapidement joue un rôle essentiel dans la démonstration du texte. Mais, une fois le théorème 12 établi, on constate aisément qu'elle est inutile; $\mu_n(\rho)$ tend en probabilité vers μ quand ρ tend vers zéro, et cela d'une manière uniforme par rapport à n .

2°. Pour démontrer que μ est presque sûrement nul, observons d'abord que sa valeur probable m est finie. Cela résulte aisément de ce que

$$Pr\{\mu > 4l^2\} < Pr\{\max_{0 \leq t \leq 1} |X(t)|, |Y(t)| > l\} \leq 4\sqrt{\frac{2}{\pi}} \int_l^\infty e^{-u^2/2} du.$$

Par suite, d'après le caractère d'homogénéité stochastique de la courbe, les mesures des arcs $A(0)A(1)$, $A(1)A(2)$, et $A(0)A(2)$, arcs que nous désignerons respectivement par C_1 , C_2 et C' , ont respectivement pour valeurs probables m , m , et $2m$. Comme on a évidemment

$$\mathcal{E}\{\text{mes } C'\} = \mathcal{E}\{\text{mes } C_1\} + \mathcal{E}\{\text{mes } C_2\} - \mathcal{E}\{\text{mes } C_1C_2\}$$

(C_1C_2 étant l'ensemble des points communs à C_1 et C_2), il en résulte que

$$\mu' = \mathcal{E}\{\text{mes } C_1C_2\} = 0,$$

c'est-à-dire que la mesure de C_1C_2 est presque sûrement nulle.

Nous allons en déduire que μ est presque sûrement nul. Observons à cet effet qu'on ne change rien à μ en supposant $A(1)$ connu et en construisant C_1 et C_2 en partant de ce point: ce sont deux déterminations indépendantes l'une de l'autre et dépendant de la même loi de probabilité. Les probabilités qu'un point M appartienne à C_1 , C_2 , et C_1C_2 sont alors $\phi(M)$, $\phi(M)$ et $\phi^2(M)$, et l'on a

$$\mu = \iint \phi(M) dx dy, \quad \mu' = \iint \phi^2(M) dx dy,$$

(les intégrations étant étendues à tout le plan); $\mu' = 0$ entraîne donc $\mu = 0$; l'un et l'autre équivalent à: $\phi(M)$ est presque partout nul.

Le théorème 12 est ainsi démontré. On remarque le rôle que joue, dans la dernière partie du raisonnement, l'indépendance stochastique de C_1 et C_2 [une fois le point $A(1)$ connu]. Il importe d'avoir ce point présent à l'esprit pour des extensions que nous indiquerons plus loin sans reprendre tout le raisonnement.

Observons d'autre part que $\phi(M)$, qui est évidemment une fonction $\psi(r)$ de la distance r du point M au point $A(1)$, est non seulement presque partout nul, mais est nul pour tout r positif. Pour le démontrer, il suffit de démontrer que c'est une fonction non croissante de r ; cela résulte évidemment de la similitude stochastique des arcs $A(0)A(1)$ et $A(1 - k^2)A(1)$; d'après cette similitude, un point situé à la distance kr du point $A(1)$ a la probabilité $\psi(r)$ d'appartenir au second de ces arcs, et par suite, si $k < 1$, une probabilité $\psi(kr) \geq \psi(r)$ d'appartenir au premier. On a donc bien $\psi(kr) \geq \psi(r)$, donc $\psi(r) = 0$ pour tout r positif.

Quant au point $A(1)$ lui-même, il est évidemment sur la courbe; mais on voit aisément que la probabilité qu'il soit double est nulle. Quoiqu'il y ait une infinité non dénombrable de points doubles, les points doubles sont sur C des points exceptionnels; leurs cotes constituent un ensemble de mesure nulle.

3°. Indiquons maintenant une condition nécessaire pour qu'une courbe remplisse un ensemble de mesure superficielle positive.

THÉORÈME 13. *Pour qu'une courbe continue remplisse un ensemble de mesure superficielle positive m , il faut que, pour des points de division convenablement choisis, mais arbitrairement denses sur la courbe, la somme*

$$(34) \quad B_n = \sum (\Delta l)^2$$

soit au moins égale à cm , c étant une constante (supérieure à $3\sqrt{3}/\pi$; nous ne connaissons pas la meilleure valeur possible pour cette constante).

Quand nous disons que les points de division sont arbitrairement denses, nous voulons dire que, quelle que soit la suite de nombres croissants entre zéro et un, $t_0 = 0, t_1, t_2, \dots, t_k = 1$, il y a au moins un des points considérés dans chacun des arcs $A(t_{h-1})A(t_h)$.

Désignons par μ_h la mesure superficielle de cet arc $A(t_{h-1})A(t_h)$. On a évidemment $\sum \mu_h \geq m$.

Or si un ensemble fermé a une mesure superficielle μ_h , on peut trouver dans cet ensemble deux points dont la distance soit au moins égale au diamètre λ_h du cercle d'aire μ_h . On vérifie en effet facilement qu'un contour dont aucune corde n'atteint cette longueur ne peut pas entourer une aire égale ou supérieure à μ_h .

Soit donc $A(t'_h)A(t''_h)$ une corde de l'arc $A(t_{h-1})A(t_h)$ de longueur au moins égale à λ_h ; on peut supposer $t'_h < t''_h$. Pour la ligne polygonale

$$A(0)A(t'_1)A(t''_1)A(t_1)A(t'_2) \dots A(t''_h) \dots A(1)$$

inscrite dans C , la somme (34) est au moins égale à

$$\sum \lambda_h^2 \geq \frac{4}{\pi} \sum \mu_h \geq \frac{4}{\pi} m,$$

ce qui établit le théorème 13, sauf en ce qui concerne la valeur de la constante.

On peut améliorer la valeur de cette constante, et en même temps obtenir un autre résultat d'un certain intérêt, en mettant en évidence trois points $A(t'_h)A(t''_h)A(t'''_h)$ de chaque arc $A(t_{h-1})A(t_h)$. On voit aisément qu'on peut toujours les choisir de manière que l'aire du triangle dont ils sont les sommets soit au moins égale à $c'\mu_h$ ($c' = 3\sqrt{3}/4\pi$); dans le cas d'une aire

circulaire, cette valeur représente l'aire du triangle équilatéral inscrit, et ne peut pas être dépassée; en dehors des cas des aires circulaires elliptiques, elle peut sûrement être dépassée.

Les triangles

$$A(t_1')A(t_1'')A(t_1'''), \quad A(t_1''')A(t_1)A(t_2'), \quad A(t_2')A(t_2'')A(t_2'''), \dots$$

forment alors une chaîne de triangles inscrits, analogue à l'aire S'_n du § 5, 1°, et dont l'aire totale est au moins $c'\Sigma\mu_n \geq c'm$. Il s'agit, bien entendu, de la somme des aires de ces triangles, prises en valeur absolue. *L'existence d'une telle chaîne pour laquelle cette somme soit au moins égale à $c'm$ est donc une condition nécessaire pour que la courbe remplisse un ensemble de mesure superficielle au moins égale à m .*

Considérons alors la ligne polygonale inscrite dans C ayant pour sommets tous ceux de ces triangles. Dans un triangle, la somme des carrés de deux côtés est au moins égale à quatre fois la surface. La somme B_n relative à cette ligne polygonale est donc au moins égale à $4c'm$; on obtient ainsi la constante $4c' = 3\sqrt{3}/\pi$ indiquée dans l'énoncé du théorème 13. Il est d'ailleurs évident qu'elle n'est pas la plus grande valeur possible pour la constante c de ce théorème. Il serait intéressant de déterminer cette valeur maxima; il ne nous a pas paru au contraire utile d'allonger les raisonnements pour obtenir une valeur un peu plus grande que $3\sqrt{3}/\pi$, mais qui ne serait pas la valeur maxima.

Indiquons d'autre part sans démonstration que, si l'on introduit le hasard dans le choix des points de division comme nous l'avons fait pour définir la notion d'oscillation brownienne, au moins pour une représentation paramétrique convenable de la courbe étudiée (la probabilité étant mesurée par la variation du paramètre), et pour les valeurs assez petites de c , on a

$$\limsup_{n \rightarrow \infty} \Pr\{B_n \geq cm^2\} = \alpha > 0.$$

Les modes de division qui réalisent la condition $B_n \geq cm^2$ n'apparaissent donc pas comme exceptionnels. La probabilité α ne peut que croître quand on prend pour c des valeurs de plus en plus petites; mais il n'est pas du tout certain qu'elle varie d'une manière continue; on peut se demander si l'on n'est pas en présence d'un de ces cas, fréquents dans la théorie des probabilités dénombrables, où la probabilité ne peut pas être comprise entre zéro et un; elle passerait brusquement d'une de ces valeurs à l'autre pour une valeur déterminée de c .

Revenant aux résultats établis d'une manière sûre, nous voyons qu'il s'en dégage deux idées. L'une c'est que: *l'aire totale des chaînes de triangles inscrits analogues à l'aire S'_n du § 5, 1° est en quelque sorte une approxima-*

tion de la mesure superficielle de la courbe ; pour que la courbe puisse remplir une aire, il faut qu'elle ne soit pas très petite. D'autre part il faut que, au moins pour une représentation paramétrique convenable, la longueur Δl de la corde $A(t)A(t+dt)$ soit en général de l'ordre de grandeur de \sqrt{dt} ou plus grande, ce qui entraîne cette conséquence indépendante de la représentation paramétrique que B_n n'est pas très petit ; si B_n n'est pas très petit, la courbe fait assez de détours infiniment petits pour pouvoir remplir une aire, et les grandes valeurs prises par B_n , pour n infini, mesurent assez bien ce qu'on pourrait appeler la possibilité pour la courbe de remplir une aire.

4°. Introduisons maintenant une nouvelle idée. Nous considérons spécialement les courbes pour lesquelles, comme pour le mouvement brownien, la corde de l'arc décrit pendant le temps dt est de l'ordre de grandeur de \sqrt{dt} , de sorte que les sommes B_n relatives à un arc fini ne sont, ni très petites, ni très grandes. La courbe fait ainsi exactement assez de détours infiniment petits pour pouvoir remplir une aire. Mais ce n'est qu'une condition nécessaire, non suffisante : pour que la courbe remplisse exactement une aire, il faut de plus une organisation de ces détours infiniment petits que le hasard n'a aucune chance de produire. Seule une loi mathématique précise peut guider le cheminement du point mobile dans des zones déjà en grande partie recouvertes de manière qu'il ne se déplace que dans les vides, et finisse par les remplir. Les courbes Γ_0 et Γ_1 dont nous parlerons au paragraphe suivant donnent des exemples de cette circonstance.

Lorsque le hasard joue un rôle suffisant, si la courbe comporte assez de détours infiniment petits pour remplir une aire m , on doit donc s'attendre à ce qu'elle remplisse seulement une aire $m' = m/k < m$, les différentes parties de cette aire étant en moyenne remplies k fois ($k > 1$) ; si k est fini, m' est positif ; si k est infini, m' est nul. D'après le théorème 12, c'est la seconde circonstance qui est réalisée. Nous allons présenter quelques remarques qui pourraient conduire à une nouvelle démonstration, mais qui, sous la forme résumée que nous leur donneront, ne sont que des raisons intuitives assez sérieuses de croire que k est infini.

A cause de la similitude stochastique des différents arcs de courbe, au lieu d'examiner un même arc à des échelles de plus en plus petites, nous pouvons examiner des arcs de plus en plus grands à une échelle déterminée. Cela nous conduit par exemple à étudier la ligne brisée $A(0)A(1)A(2) \cdots A(n) \cdots$ indéfiniment prolongée ; nous supposons les sommets de cette ligne marqués sur une feuille de papier quadrillé, les côtés des carrés du quadrillage étant égaux à l'unité de longueur ou un peu plus petits que cette unité, de manière que deux points consécutifs $A(n)$ et $A(n+1)$ aient des chances appréciables

de ne pas être dans le même carré. Montrons d'abord qu'il y a une probabilité unité pour que le carré du quadrillage qui contient $A(0)$ contienne une infinité d'autres sommets de cette ligne brisée.

Il suffit à cet effet de montrer que, quel que soit $A(n)$ supposé connu, on peut déterminer N de manière que l'un au moins des points $A(n+1)$, $A(n+2)$, \dots , $A(n+N)$ soit dans le carré du quadrillage qui contient $A(0)$, et cela dans des cas dont la probabilité ne soit pas très petite. En effet, sauf dans des cas peu probables, ces N points sont à une distance de $A(n)$ ne dépassant pas $c\sqrt{N}$, c étant une constante convenablement déterminée. Ils ne peuvent donc se répartir qu'entre des carrés du quadrillage dont le nombre ne dépasse pas $\pi(c\sqrt{N} + \sqrt{2})^2$, soit sensiblement $\pi c^2 N$, et, pour chacun de ces carrés on peut borner inférieurement la probabilité qu'il contienne un de ces points (comme il s'agit de principes bien connus, nous n'insistons pas sur les détails de la démonstration). Il suffit alors que $c\sqrt{N}$ dépasse la distance $A(0)A(n)$ pour que cette conclusion s'applique au carré du quadrillage contenant $A(0)$; il y aura une probabilité supérieure à un nombre fixe α qu'il contienne un point $A(v)$ d'indice compris entre n et $n+N$.

On peut alors sûrement déterminer une suite d'entiers croissants $n_1, n_2, \dots, n_h, \dots$, tels que, une fois les n_h premiers points $A(v)$ connus, il y ait une probabilité supérieure à α qu'un des $N_h = n_{h+1} - n_h$ suivants soit dans le carré du quadrillage qui contient $A(0)$; N_h , dépendant du point $A(n_h)$, est aléatoire, mais sûrement borné. On sait que, dans ces conditions, il est presque sûr que l'on obtiendra indéfiniment des points $A(v)$ situés dans le carré qui contient $A(0)$.¹⁰

Le même résultat s'applique naturellement à n'importe quel carré du quadrillage, et l'on en conclut aisément que l'ensemble des points $A(n)$ forme presque sûrement un ensemble partout dense dans le plan; il en est de même, a fortiori, de la ligne polygonale ayant ces points pour sommets, et de la courbe C elle-même.

On peut alors se représenter de la manière suivante l'aspect de cette ligne polygonale limitée à ses n premiers côtés, n étant grand. La plus grande distance de deux de ses points sera de l'ordre de grandeur de \sqrt{n} , et elle ne recouvrira certainement pas avec une grande densité la plus petite région convexe qui l'entoure; il y aura des vides, et il y aura des parties de cette région R où la ligne considérée ne passe qu'une fois ou un petit nombre de fois. Mais le fait que les remarques précédentes s'appliquent à n'importe quel carré

¹⁰ Les remarques que nous venons d'exposer dans les deux derniers alinéas reproduisent à peu près des considérations exposées par M. G. Pólya dans une conférence faite au Colloque sur les principes du calcul des probabilités tenu à Genève en octobre 1937 et publiée chez Hermann.

du quadrillage intérieur à cette région et dont on sait qu'il contient au moins un sommet $A(\nu)$,¹⁷ prouve que la plupart des carrés qui contiennent un sommet en contiennent un grand nombre. Si alors, pour avoir une idée de l'aire recouverte, on considère, soit l'ensemble des carrés du quadrillage contenant au moins un sommet, soit la chaîne des triangles $A(2\nu)A(2\nu+1)A(2\nu+2)$, on voit que l'aire ainsi définie ne sera pas, comme on pourrait s'y attendre à première vue, de l'ordre de grandeur de n ; elle sera petite par rapport à n , et composée en grande partie de régions recouvertes un grand nombre de fois. Il en résulte nécessairement qu'il y a des vides, dont les plus grands seront une fraction non négligeable de la région R ,¹⁸ et qui seraient seulement recouverts par le prolongement de la ligne polygonale étudiée au delà de ses n premiers côtés.

Utilisons maintenant la similitude stochastique des différents arcs de la courbe C . Les résultats précédents peuvent s'appliquer à l'étude de la ligne

$$A(0)A(1/n)A(2/n) \cdots A(1),$$

pour une valeur très grande de n . Nous trouvons d'abord un résultat qui rejoint les remarques du § 3, 5°, *in fine*: la courbe n'atteint un point, en général, qu'après avoir passé près de lui un grand nombre de fois; la distance $A(t)A(t+\tau)$, quand τ tend vers zéro, est en général de l'ordre de grandeur de $\sqrt{\tau}$; mais elle est parfois plus grande et parfois plus petite, ce qui donne à la courbe l'aspect d'une succession de bouches de plus en plus petites et de plus en plus voisines du point $A(t)$; à une échelle excessivement petite, on pourra voir le point $A(t+\tau)$ s'approcher de $A(t)$, puis s'en éloigner, et cela un grand nombre de fois avant que la distance $A(t)A(t+\tau)$ cesse d'être appréciable.

D'autre part, en ce qui concerne l'aire, nous voyons qu'une chaîne de triangles inscrits comme celle désignée au § 5 par S'_n , bien que la somme des aires de ces triangles prises en valeur absolue ait pour n infini une limite positive, ne recouvre qu'une aire de plus en plus petite, mais recouverte un nombre de fois de plus en plus grand. Cette aire pouvant être considérée comme une approximation de celle recouverte par la courbe, on est conduit à conclure que la mesure superficielle de la courbe est nulle. Une extension convenable du théorème 13 permettrait de rendre ce raisonnement rigoureux,

¹⁷ La nécessité de cette restriction est évidente: un point pris au hasard dans R est à une distance de $A(0)$ qui est de l'ordre de grandeur de \sqrt{n} . Ce n'est donc qu'après avoir placé un nombre de sommets $A(\nu)$ grand par rapport à n qu'on a une grande probabilité d'en trouver un qui soit voisin du point donné.

¹⁸ Il faut bien en effet que pour un point pris au hasard dans les régions vides on puisse appliquer le raisonnement de la note précédente. Or on ne le peut pas pour un point qui serait à une distance d'un des $A(\nu)$ petite par rapport à \sqrt{n} .

et conduirait à une nouvelle démonstration du théorème 12. La démonstration initiale est évidemment plus simple; mais les remarques qui précèdent nous ont paru utiles pour montrer que: *la courbe C , tout en comportant assez de détours infiniment petits pour recouvrir une aire, a cependant une mesure superficielle nulle, parce que l'allure désordonnée du point mobile ne permet pas le balayage méthodique d'une aire; il est infiniment peu probable que ce balayage soit réalisé.*

7. Généralisations diverses. 1°. Un des caractères essentiels du mouvement brownien est la similitude stochastique de deux arcs quelconques de trajectoire. Ce caractère est indépendant du nombre de dimensions de l'espace considéré, et subsiste par une transformation affine, c'est-à-dire qu'à la loi de Gauss isotrope on peut substituer la loi de Gauss non isotrope. Mais les lois stables autres que celles de Gauss conduisant à des courbes presque sûrement discontinues, il ne semble pas que l'on puisse trouver d'autres schémas présentant ce caractère et conduisant à des courbes continues.¹⁹

Par contre il est facile de définir des schémas très variés pour lesquels la courbe décrite quand t varie de zéro à un est une réunion d'arcs stochastiquement semblables à la courbe entière. Pour nous limiter, nous n'étudierons que les courbes pour lesquelles les deux arcs $A(0)A(\frac{1}{2})$ et $A(\frac{1}{2})A(1)$ sont stochastiquement semblables à la courbe entière; chacun de ces arcs se décomposant à son tour dans les mêmes conditions, et ainsi de suite, nous voyons que chacun des arcs $A(h \cdot 2^{-n})A[(h+1)2^{-n}]$ est stochastiquement semblable à la courbe entière. Les points dont les cotes sont de la forme $h \cdot 2^{-n}$ sont alors des points particuliers de la courbe; l'allure de la courbe en un tel point ne ressemblera pas à son allure en un point quelconque. Les lignes polygonales L'_n , ayant ces points pour sommets, et les chaînes de triangles inscrits désignées par S'_n au début du § 5 se distingueront essentiellement des autres lignes polygonales inscrites et des autres chaînes de triangles inscrits; on doit s'attendre à trouver pour les L'_n et les S'_n des propriétés simples non susceptibles d'être étendues sans modification aux autres lignes inscrites L_n et aux autres chaînes de triangles inscrits.

D'autre part, ce qui n'était pas possible (en dehors du cas du mouvement rectiligne et uniforme) lorsqu'on exigeait la similitude de n'importe quel arc de courbe avec la courbe entière, devient ici possible: il peut s'agir de similitude véritable, et non de similitude stochastique. On retrouve ainsi des courbes dont nous avons fait une étude systématique dans un mémoire récent (*Journal de l'Ecole Polytechnique*, 1938); deux de ces courbes seront con-

¹⁹ En tout cas cela est évident si l'on se borne aux schémas pour lesquels les déplacements successifs du point mobile sont stochastiquement indépendants.

sidérées dans la suite, et désignées par Γ_0 et Γ_1 ; Γ_1 est la courbe bien connue qui remplit l'aire d'un triangle rectangle isocèle; pour la courbe Γ_0 , nous renvoyons à notre mémoire de 1938 pour la démonstration de ses principales propriétés.

2°. Nous allons considérer en premier lieu les courbes Γ pour lesquelles le triangle $A(0)A(\frac{1}{2})A(1)$ est un triangle rectangle isocèle dont $A(0)A(1)$ est l'hypoténuse. Chacun des 2^n triangles qui constituent l'aire S'_n sera aussi un triangle rectangle isocèle; si l'on prend $A(0)A(\frac{1}{2})$ pour unité de longueur, les côtés de l'angle droit de chacun de ces triangles auront la longueur q^n ($q = 1/\sqrt{2}$). L'hypoténuse étant placée, le sommet de l'angle droit a deux positions possibles, et l'aire du triangle, égale en valeur absolue à $2^{-(n+1)}$ pour les triangles de S'_n , sera positive ou négative suivant le sommet choisi. La courbe sera donc bien définie par la donnée d'une succession de signes; nous désignerons par $\epsilon_n^{(h)} = \pm 1$ le signe lié au $h^{\text{ième}}$ triangle de S'_n . Nous supposons $\epsilon_0^{(1)} = 1$, ce qui n'est pas une restriction essentielle.

Nous étudierons spécialement les deux courbes non aléatoires Γ_0 et Γ_1 définies, la première par $\epsilon_n^{(h)} = 1$, la seconde par $\epsilon_n^{(h)} = (-1)^n$, et les deux courbes aléatoires Γ_2 et Γ_3 pour chacune desquelles les signes seront déterminés par des tirages au sort à chances égales pour les deux signes; mais pour Γ_2 le signe ne dépendra que de n et un même tirage au sort déterminera l'orientation de tous les triangles de S'_n ; pour Γ_3 il y aura un tirage au sort pour chaque triangle.

Bien entendu, des règles quelconques ne donneraient pas des courbes composées d'arcs stochastiquement semblables à la courbe entière. L'énumération complète des courbes Γ pour lesquelles il y a similitude (effective, ou stochastique) entre chacun des arcs $A(0)A(\frac{1}{2})$ et $A(\frac{1}{2})A(1)$ et la courbe entière serait assez longue. Il existe en outre des courbes Γ composées de quatre (ou huit, ou seize) arcs stochastiquement semblables à la courbe, et non deux; tel serait le cas si l'on admet qu'un même tirage au sort détermine l'orientation des triangles de S'_n pour deux (ou trois, ou quatre) valeurs consécutives de n .

Pour la courbe Γ_0 , l'aire totale des triangles de S'_n , qui sont tous orientés positivement, a la valeur $\frac{1}{2}$, aire du triangle initial. Les différents triangles de S'_n ne se recouvrent jamais, de sorte que l'aire totale de S'_n est $\frac{1}{2}$. Cela conduit à penser que la courbe Γ_0 recouvre une aire égale à $\frac{1}{2}$; c'est ce que nous avons démontré dans le travail cité tout à l'heure.

D'autre part l'aire $S'_0 + S'_1 + \dots + S'_{n-1}$, aire comprise entre L'_0 et L'_n comptée en affectant chacune de ses parties d'un coefficient numérique qui indique combien de fois elle est entourée, est égale à $n/2$; elle augmente indéfiniment, et l'aire comprise entre la courbe Γ_0 et sa corde est infinie.

On se l'explique bien en observant que la courbe est composée de boucles qui tournent toujours dans le même sens. On en déduit aisément que, pour un choix quelconque des points de division, on aurait toujours une aire infinie.

Dans le cas de la courbe Γ_1 , chacune des aires S'_n recouvre exactement l'aire du triangle initial; à la limite, la courbe remplit le triangle: pour Γ_0 et Γ_1 , une loi mathématique précise, pour des raisons évidentes dans le cas de Γ_1 et beaucoup plus cachées dans le cas de Γ_0 , réussit à faire ce que le hasard ne peut pas faire: la courbe remplit une aire sans qu'aucune partie de cette aire soit recouverte plus d'une fois.

En tenant maintenant compte des signes, l'aire comprise entre L'_0 et L'_n se présente sous la forme

$$\frac{1}{2} - \frac{1}{2} + \frac{1}{2} - \frac{1}{2} + \dots,$$

de sorte qu'elle est égale à zéro si n est pair et à $\frac{1}{2}$ si n est impair. On se rend bien compte de ce fait, géométriquement, en observant qu'après suppression de segments rectilignes dont chacun est parcouru une fois dans chaque sens, les lignes L'_n se réduisent à L'_0 ou à L'_1 , suivant la parité de n . L'aire comprise entre la courbe Γ_1 et le segment initial $A(0)A(\frac{1}{2})$ apparaît ainsi comme indéterminée entre zéro et un.

Pour la courbe Γ_2 , l'aire S'_n a la valeur $\epsilon_n/2$. L'aire comprise entre la courbe et le segment initial $A(0)A(1)$ est alors comparable au gain d'un joueur dans une partie de pile ou face indéfiniment prolongée; elle est indéterminée, non entre deux limites fixes, mais entre $-\infty$ et $+\infty$. D'autre part un raisonnement identique à celui fait à propos de Γ_2 dans notre mémoire citée ci-dessus permet de montrer que les différents triangles d'une même aire S'_n ne se recouvrent pas: si l'on part d'un réseau de triangles recouvrant le plan, chaque succession de signes $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ conduit à un réseau d'aires S'_n recouvrant exactement le plan une fois et une seule, et, à la limite, on obtient un réseau de courbes Γ_3 , infiniment enchevêtrées les unes dans les autres, mais recouvrant le plan une fois et une seule; il y a lieu de penser que chacune recouvre une aire égale à celle du triangle initial.²⁰

Le fait que le même tirage au sort définisse les orientations de tous les triangles d'une même aire S'_n suffit à constituer cette loi précise qui fait ce que le hasard ne saurait faire: deux triangles de S'_n ne peuvent pas se recouvrir.

Il n'en est plus de même pour la courbe Γ_3 , dans la définition de laquelle le hasard joue un rôle beaucoup plus grand. D'abord chaque aire S'_n , compte tenu des signes de ses triangles, est assimilable au gain d'un joueur après 2^n coups de pile ou face, l'enjeu à chaque coup étant $2^{-(n+1)}$. C'est une

²⁰ On démontre du moins aisément que chacune a une mesure superficielle au moins égale à celle de ce triangle.

variable asymptotiquement gaussienne, dont l'écart quadratique moyen est q^{n+2} ($q = 1/\sqrt{2}$). La série $\sum S'_n$ est donc une série à termes indépendantes, qui converge en moyenne quadratique, donc presque sûrement. L'aire S comprise entre la courbe et sa corde est donc stochastiquement bien définie, dans les mêmes conditions que pour le mouvement brownien; on peut aussi montrer qu'avec des points de division choisis au hasard il y a, dans les mêmes conditions que pour le mouvement brownien, convergence presque sûre vers la même limite S ; mais il ne s'agit pas d'une aire définie au sens de Riemann.

Comme c'est une somme de termes aléatoires indépendants, on définit facilement, par sa fonction caractéristique, la loi dont elle dépend. Cette fonction caractéristique est

$$(55) \quad \prod_0^\infty \left(\cos \frac{z}{2^{n+1}} \right)^{2^n} = \frac{\sin z}{z} \prod_0^\infty \left(\frac{2^{n+1}}{z} \sin \frac{z}{2^{n+1}} \right)^{2^n}.$$

La deuxième expression, correspondant à un groupement évident des facteurs de la première, donc aussi au groupement correspondant des triangles dont S est la somme, montre que S est la somme de variables indépendantes ayant chacune une fonction caractéristique de la forme $(\lambda/z) \sin z/\lambda$, c'est-à-dire que cette variable est choisie arbitrairement entre $-\lambda$ et $+\lambda$ avec une répartition uniforme de la probabilité. Elle dépend ainsi d'une loi absolument continue, et il en est de même de S .

Montrons maintenant que: *la mesure superficielle de la courbe Γ_3 est nulle*. Le principe du raisonnement est le même que dans le cas du mouvement brownien (§ 6, 1° et 2°). Mais ici, au lieu d'un facteur $\phi(M)$ qui intervient deux fois, il faut introduire deux facteurs $\phi_1(M)$ et $\phi_2(M)$ qui représentent respectivement les probabilités que M appartienne aux arcs $A(0)A(\frac{1}{2})$ et $A(\frac{1}{2})A(1)$; ils sont respectivement égaux à $\psi(r, \theta)$ et $\psi(r, \pi/2 - \theta)$, r désignant la distance $A(\frac{1}{2})M$ et θ l'angle $A(0)A(\frac{1}{2})M$. On sait que le produit $\psi(r, \theta)\psi(r, \pi/2 - \theta)$ est presque partout nul; il s'agit de montrer que chacun des facteurs est presque partout nul. Ce qui était évident lorsque les deux facteurs étaient égaux ne l'est plus ici.

Mais un artifice très simple va nous permettre d'arriver au résultat. Il y a une chance sur quatre pour que $\epsilon_1^{(1)} = \epsilon_1^{(2)} = -1$; dans ce cas les points $A(\frac{1}{4})$ et $A(\frac{3}{4})$ coïncident, et les arcs $A(\frac{1}{4})A(\frac{1}{2})$ et $A(\frac{1}{2})A(\frac{3}{4})$ sont deux déterminations indépendantes d'une même courbe aléatoire, et la probabilité qu'un point M appartienne à l'un ou à l'autre de ces arcs a une même valeur $\phi(M)$. Si alors Γ_3 avait une mesure superficielle positive, et cela dans des cas de probabilité positive, il y aurait aussi une probabilité positive que les arcs $A(\frac{1}{4})A(\frac{1}{2})$ et $A(\frac{1}{2})A(\frac{3}{4})$, stochastiquement semblables à Γ_3 , aient des mesures superficielles positives, et que de plus $A(\frac{1}{4})$ et $A(\frac{3}{4})$ coïncident. L'indépendance de ces arcs, une fois les points $A(\frac{1}{4})$, $A(\frac{1}{2})$ et $A(\frac{3}{4})$ placés, permet

de terminer le raisonnement presque comme dans le cas du mouvement brownien: la mesure de l'ensemble des points communs aux deux arcs considérés pourrait être positive, dans des cas de probabilité positive. Il en serait de même des points communs aux arcs $A(0)A(\frac{1}{2})$ et $A(\frac{1}{2})A(1)$. Or, par la première partie du raisonnement, qui subsiste sans modification, on sait que c'est impossible; ce qui établit le résultat annoncé.

On voit que, si ce résultat a pu être obtenu, c'est parce que la part du hasard est bien plus grande que pour la courbe Γ_2 ; pour cette courbe, les arcs $A(\frac{1}{4})A(\frac{1}{2})$ et $A(\frac{1}{2})A(\frac{3}{4})$ sont égaux; pour Γ_3 , ils sont stochastiquement indépendants [une fois le point $A(\frac{1}{2})$ placé]; cette indépendance joue un rôle essentiel dans le raisonnement qui précède.

3°. Pour terminer l'étude des courbes Γ , nous allons présenter quelques remarques relatives à la somme

$$(34) \quad B_n = \sum (\Delta l)^2,$$

qui, étendue aux côtés d'une des lignes L'_n , a la valeur non aléatoire 2. Si on l'étudie dans le cas d'une ligne L_n dont les sommets sont choisis au hasard entre zéro et un, des considérations analogues à celles exposées à propos du mouvement brownien montrent que, si les points de division, une fois choisis, sont conservés, il est presque sûr que pour n infini, B_n est infiniment peu différent de sa valeur probable.²¹ Mais cette valeur probable n'est plus une constante; elle est de la forme $P(\log n) + \epsilon_n$, $P(\lambda)$ étant une fonction périodique, et ϵ_n tendant vers zéro. La suite des B_n présentera donc, sur l'échelle logarithmique, des oscillations asymptotiquement périodiques.

Pour établir ce résultat, considérons d'abord la valeur probable $\mu^2 = \phi^2(\tau)$ de $(\Delta l)^2$, Δl étant la longueur d'une corde pour laquelle Δt a une valeur donnée τ , et la cote t de son origine étant choisie au hasard entre 0 et $1 - t$. Si τ devient deux fois plus petit, μ^2 devient à peu près deux fois plus petit; on obtient évidemment $\phi^2(\tau)/2$ comme valeur probable de $(\Delta l)^2$ pour $\Delta t = \tau/2$ et t choisi au hasard entre 0 et $(1 - \tau)/2$ ou entre $\frac{1}{2}$ et $1 - \tau/2$. La valeur probable de $(\Delta l)^2$, pour $\Delta t = \tau/2$ et t choisi au hasard entre 0 et $1 - \tau/2$ sera donc de la forme $\phi^2(\tau)/2 [1 + O(\tau)]$ {on le voit aisément en observant que Δl est toujours $O[\phi(\tau)]$ }. Si l'on donne alors successivement à τ les valeurs $\tau, \tau/2, \tau/4, \dots$, on voit que $(2^p/\tau)\phi^2(\tau/2^p)$ tend vers une limite, pour p infini, ce qui revient à dire que

$$(56) \quad \frac{\phi^2(\tau)}{\tau} = P_1(\log \tau) + \epsilon(\tau),$$

$\epsilon(\tau)$ tendant vers zéro avec τ , et P_1 ayant la période $\log 2$.

²¹ Pour les schémas aléatoires Γ_2 et Γ_3 , il est bien entendu qu'il s'agit de la valeur

Considérons maintenant les points t_1, t_2, \dots, t_{n-1} choisis au hasard entre zéro et un, qui sont les cotes de sommets de L_n . On sait que chacun des intervalles $\Delta t = \tau$ séparés par ces points dépend de la loi définie par

$$Pr\{n\tau > x\} = \left(1 - \frac{x}{n}\right)^n \rightarrow e^{-x} \quad (n \rightarrow \infty),$$

et que, si n est grand, les différentes valeurs possibles de $n\tau$ sont réalisées avec des fréquences très probablement très peu différentes de leurs probabilités (on peut même préciser ce résultat au sens de la loi forte des grands nombres). Il y a d'ailleurs, asymptotiquement, indépendance stochastique entre l'origine t et la longueur τ des intervalles considérés, de sorte que, pour un intervalle de longueur τ connue, la valeur probable de $(\Delta l)^2$ est bien $\phi^2(\tau)$. On en déduit que l'on a asymptotiquement

$$\begin{aligned} \mathcal{E}\{B_n\} &= \int_0^\infty P_1(\log \tau) e^{-n\tau} n^2 \tau d\tau + \epsilon_n \\ &= \int_0^\infty P_1\left(\log \frac{x}{n}\right) e^{-x} dx + \epsilon_n = P(\log n) + \epsilon_n, \end{aligned}$$

ϵ_n tendant vers zéro et $P(\log n)$ étant une fonction périodique de période $\log 2$, c. q. f. d.

On remarque que $P(\log n)$, étant une moyenne entre les différentes valeurs de $P_1(\log x/n)$, ne varie qu'entre des limites assez voisines l'une de l'autre. Comme B_n , si n est grand, diffère très probablement (même presque sûrement très peu) de sa valeur probable, on ne peut pas parler d'une oscillation brownienne bien définie comme dans le cas du mouvement brownien, mais cette oscillation est indéterminée entre deux limites voisines l'une de l'autre; cette indétermination n'a d'ailleurs pas un caractère aléatoire: B_n diffère en effet très probablement (et même presque sûrement, dans les mêmes conditions que pour le mouvement brownien) de la fonction non aléatoire $P(\log n)$.²²

On peut d'ailleurs échapper à ces oscillations périodiques en modifiant le choix des points de division de la manière suivante: nous choisirons un point de division t_0 au hasard entre zéro et un, avec répartition uniforme de la probabilité; puis deux points $t_1^{(1)}$ et $t_1^{(2)}$ respectivement dans les deux intervalles $(0, t_0)$ et $(t_0, 1)$; puis quatre nouveaux points dans les intervalles ainsi

probable a priori $\mathcal{E}\{B_n\}$, et que c'est en tenant compte à la fois du choix de la courbe et de celui des t_ν que nous disons que $B_n - \mathcal{E}\{B_n\}$ tend presque sûrement vers zéro.

²² C'est donc par erreur que, dans ma Note du 12 décembre 1938, j'avais indiqué les courbes Γ_0 et Γ_1 (désignées dans cette Note par C_0 et C_1) comme modèles de mouvement brownien. Du moins il semble que ce soit une erreur. Il n'y a première vue aucune raison de penser que la fonction périodique $P_1(\log \tau)$ se réduise à une constante; mais je n'ai pas démontré que cette hypothèse est exclue. On remarque d'ailleurs qu'il est a priori possible qu'elle soit constante pour Γ_0 et variable pour Γ_1 , ou inversement.

distingués, et ainsi de suite. Après p opérations analogues, on aura défini une ligne polygonal L_p'' à 2^p côtés, inscrite dans Γ . On peut penser que l'oscillation périodique signalée ci-dessus disparaît ici simplement parce qu'on ne considère que des valeurs entières de $p = \log n / \log 2$; mais il se produit aussi une autre circonstance remarquable. Les $n = 2^p$ valeurs de $\tau = \Delta t$ correspondant aux côtés de L_p'' ne sont pas ici pour la plupart de l'ordre de grandeur de $1/n$; les n valeurs des produits $n\tau$ se répartissent sur un intervalle beaucoup plus étendu que dans le cas précédent, et, pour n'importe quel intervalle fini sur l'échelle des $\log 1/\tau$, la probabilité tend vers zéro pour n infini et tend à s'y répartir avec une densité constante. On aura alors à considérer, au lieu de $P(\log n)$, une moyenne entre les différentes valeurs de $P(\log 1/\tau)$ qui se réduira à la limite à

$$B = \frac{1}{\log 2} \int_0^{\log 2} P_1(u) du,$$

et les valeurs de $B_n = B_p''$ correspondant aux lignes polygonales L_p'' tendent en probabilité, et même presque sûrement, vers B . Il faut remarquer qu'il n'y a aucune raison de penser que B a la même valeur 2 que dans le cas des lignes L_p' ; c'était une valeur particulière tenant au rôle particulier qui jouent les lignes L_p' dans la définition des lignes Γ ; ici il s'agit d'une moyenne, presque sûrement réalisée dans les conditions où nous nous sommes placés. On aurait d'ailleurs la même valeur limite pour B_n si l'on partait d'une division initiale de l'intervalle $(0, 1)$ en h intervalles égaux (ou choisis au hasard), dont chacun serait subdivisé ensuite comme il vient d'être indiqué. Dans les remarques qui précèdent, on pourrait s'attendre à trouver comme limite, au lieu de la constante B , une fonction périodique de $\log h$. Il n'en est rien, et cette constante B semble donner, pour chacun des types de courbes Γ , une bonne mesure de ce qu'on peut appeler l'*oscillation brownienne généralisée*; c'est une limite généralisée, ou limite en moyenne par rapport à la variable $\log n$, de la suite des B_n obtenus par le premier des processus indiqués.

Des considérations analogues, dans le cas de la courbe Γ_1 , peuvent s'appliquer à l'aire comprise entre la courbe et sa corde; on peut définir une aire stochastique généralisée qui serait nécessairement égale à la moitié de l'aire du triangle initial.²³ Dans le cas de la courbe Γ_2 , il y a presque sûrement une

²³ Il faut remarquer que nous n'avons pas exclu l'hypothèse qu'il y ait une aire stochastique non généralisée. Si c'était le cas, cela n'empêcherait pas que pour des lignes polygonales inscrites L_n convenablement choisies l'aire comprise entre L_n et $A(0)A(1)$, ne convergerait pas vers cette aire stochastique, et rien n'empêche de penser que les lignes L_n soient précisément de telles lignes exceptionnelles. Disons seulement, en répétant une idée exprimée dans la note précédente, que le mode de définition de la courbe implique la périodicité sur l'échelle logarithmique, et que nous ne voyons aucune

aire stochastique généralisée, mais variable avec cette courbe (tandis que l'oscillation brownienne généralisée ne dépend pas du choix de la courbe).

4°. Etudions maintenant les courbes obtenues en prenant pour $A(0)A(\frac{1}{2})A(1)$ un triangle isocèle de base $A(0)A(1)$ et d'angle au sommet α ; nous les désignerons par $\Gamma(\alpha)$; pour $\alpha = \pi/2$ elles se réduisent à celles que nous venons d'étudier. Nous désignerons par $\Gamma_h(\alpha)$ ($h = 0, 1, 2, 3$) la courbe pour laquelle l'orientation des triangles des aires S'_n est définie comme pour la courbe Γ_h .

Le rapport de similitude (effectif, ou stochastique) de chacun des arcs $A(0)A(\frac{1}{2})$ et $A(\frac{1}{2})A(1)$ et de la courbe entière est $q = \frac{1}{2 \sin \alpha/2}$. Si $\alpha > \pi/2$, on a $q^2 < \frac{1}{2}$; il en résulte immédiatement que, si τ est très petit, la longueur des cordes $A(t)A(t + \tau)$ est $o(\sqrt{\tau})$; l'oscillation brownienne est nulle. La courbe a alors une mesure superficielle nulle. D'autre part l'aire comprise entre la courbe et sa corde est bien définie, au sens de l'analyse ordinaire. Il est inutile d'insister davantage sur ce cas simple; le cas où $\alpha < \pi/2$, donc $q^2 > \frac{1}{2}$, est moins simple. Il faut bien entendu, pour que la suite des lignes L'_n successivement définies convergent vers une courbe, que l'on ait $q < 1$. Nous supposons donc maintenant α compris entre $\pi/2$ et $\pi/3$, et étudierons la courbe $\Gamma_3(\alpha)$ pour laquelle l'orientation de chacun des triangles de chaque aire S'_n dépend d'un tirage au sort indépendant des autres.

L'aire d'un triangle de S'_n est $\pm q^{2n}s$ (s étant l'aire du triangle initial). L'aire totale de S'_n , compte tenu des signes, a pour valeur quadratique moyenne $2^{n/2}q^{2n}s$. La condition pour que la série $\Sigma S'_n$, qui définit l'aire S , soit convergente en moyenne quadratique, et par suite presque sûrement convergente, est donc $2q^4 < 1$, c'est-à-dire $\alpha > \alpha'$, α' étant l'angle compris entre $\pi/3$ et $\pi/2$ pour lequel $8 \sin^4 \alpha'/2 = 1$. Pour ces valeurs de α , l'aire S est stochastiquement définie; pour $\alpha \leq \alpha'$, la série $\Sigma S'_n$ est essentiellement divergente, et l'on ne peut pas, même par des procédés de moyennes, définir S .

Au point de vue de la mesure superficielle de la courbe $\Gamma_3(\alpha)$, les considérations exposées à propos de Γ_3 subsistent en ce sens que, si n est grand, les portions du plan recouvertes par S'_n ont chance de l'être un grand nombre de fois. Mais en même temps la somme des aires des triangles de S'_n , prises en valeur absolue, augmente proportionnellement à $(2q^2)^n$; la courbe fait donc d'autant plus de détours infiniment petits, a d'autant plus de chances de pouvoir remplir une aire, que α est plus petit. On est donc en présence de deux causes agissant en sens contraire, et l'on ne sait pas à première vue laquelle l'emporte. On peut seulement observer qu'une des causes varie avec

raison de penser que les oscillations que cette périodicité laisse prévoir n'existent pas effectivement dans l'étude de l'aire. Le calcul d'une moyenne sur un intervalle assez étendu les fait en tout cas disparaître.

α , et cela ne semble pas être le cas pour l'autre. D'autre part la probabilité que la mesure superficielle de $\Gamma_3(\alpha)$ soit positive, n'étant pas modifiée par le résultat d'un nombre fini d'épreuves, ne peut être que zéro ou un. Il y a donc lieu de penser qu'il existe un nombre α'' (peut-être égal à α') tel que cette probabilité soit nulle pour $\alpha > \alpha''$ (ou peut-être $\alpha \geq \alpha''$) et égale à l'unité dans le cas contraire.

5°. Etudions maintenant un exemple de schéma dans lequel le rapport de similitude stochastique de chacun des arcs $A(0)A(\frac{1}{2})$ et $A(\frac{1}{2})A(1)$ sera aléatoire. Nous prendrons à cet effet pour $A(\frac{1}{2})$ un point choisi au hasard sur la circonférence de diamètre $A(0)A(1)$, ou sur l'une ou l'autre des demi-circonférences limitées à ce diamètre; de même chaque triangle de chaque aire S'_n sera un triangle rectangle ayant pour hypoténuse le côté de L'_n qui lui sert de base. Pour mettre l'orientation de ces triangles en évidence, nous supposerons dans tous les cas qu'on choisisse le sommet indéterminé sur la demi-circonférence située à gauche de ce diamètre (les sommets de L'_n étant parcourus dans le sens des t croissants), et toujours avec une répartition uniforme de la probabilité sur cette demi-circonférence (on pourrait d'ailleurs adopter d'autres règles). On conservera le point choisi, ou bien on le remplacera par le point symétrique, situé sur l'autre demi-circonférence, suivant le signe d'un nombre $\epsilon_n^{(h)}$ qui sera déterminé comme pour les courbes Γ . Nous désignerons par Γ' les courbes ainsi obtenues, et par $\Gamma'_0, \Gamma'_1, \Gamma'_2, \Gamma'_3$ les courbes, toutes aléatoires, qui correspondent respectivement aux courbes $\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4$.

On remarque que, dans tous les cas, la somme $B_n = \Sigma(\Delta l)^2$ étendue aux lignes L'_n , est égale au carré de la longueur $A(0)A(1)$, carré que nous supposons toujours égal à 2. Au point de vue de l'oscillation brownienne, nous pouvons répéter ce qui a été dit pour les courbes Γ . Si l'on choisit des points de division au hasard, B_n est indéterminé entre deux limites positives; mais le fait que, pour les lignes L'_n , on ait $B_n = 2$, suffit à montrer que la courbe fait juste assez de détours infiniment petits pour pouvoir remplir une aire, si son tracé était guidé par d'autres lois que celles du hasard.

Évaluons d'abord l'aire S'_n . Un triangle de cette aire, si son hypoténuse est Δl , a pour aire $\frac{1}{4}(\Delta l)^2 \sin \Phi$, Φ étant un angle choisi au hasard entre 0 et π ; sa valeur probable, sans tenir compte des signes, est donc $(\Delta l)^2/2\pi$, et, pour l'ensemble des triangles de S'_n , comme $B_n = 2$, la somme de ces valeurs probables est $1/\pi$. Il s'agit d'ailleurs d'une somme de termes tous très petits; on vérifie aisément qu'il y a convergence en moyenne quadratique, et même convergence presque sûre, vers cette valeur probable.

Pour les courbes Γ'_0, Γ'_1 et Γ'_2 , on se trouve alors dans les mêmes conditions que pour Γ_0, Γ_1 , et Γ_2 : chaque aire S'_n étant une somme de triangles ayant la même orientation, la série $\Sigma S'_n$ qui définit S est asymptotiquement

de la forme $\Sigma \pm 1/\pi$; elle est divergente, et l'aire S n'est pas stochastiquement définie; du moins, à première vue, elle ne semble pas l'être.

Pour la courbe Γ'_3 , au contraire, les aires des triangles de chaque S'_n ont des signes variables. La valeur probable de S'_n est nulle, et celle de S'^2_n est

$$\begin{aligned}\mathcal{E}\{(S'_n)^2\} &= \frac{1}{16} \Sigma \mathcal{E}\{(\Delta l)^4 \sin^2 \phi\} = \frac{1}{32} \Sigma \mathcal{E}\{(\Delta l)^4\} \\ &\leq \frac{1}{32} \Sigma (\Delta l)^2 \mathcal{E}\{\text{Max}(\Delta l)^2\} = \frac{1}{16} \mathcal{E}\{\text{Max}(\Delta l)^2\}.\end{aligned}$$

Cette expression est le terme général d'une série convergente. D'autre part, quand tous les sommets de la ligne L'_n (et par suite $S'_1, S'_2, \dots, S'_{n-1}$) sont connus, l'aire S'_n dépend d'une loi symétrique. On sait que ces conditions entraînent à la fois la convergence en moyenne quadratique et la convergence presque sûre de la série $S = \Sigma S'_n$; l'aire S est ainsi presque sûrement définie.

6°. Démontrons maintenant que: *les courbes Γ' ont presque sûrement une mesure superficielle nulle.* Le raisonnement qui suit, en grande partie identique à ceux faits pour le mouvement brownien et pour la courbe Γ_3 , s'applique indifféremment aux différents types de courbes Γ' .

Les courbes Γ' étant dans une région bornée, leur mesure superficielle est bornée; elle a donc une valeur probable m , positive ou nulle, mais finie. Si $A(0)$, $A(\frac{1}{2})$ et $A(1)$ sont connus, les valeurs probables des mesures superficielles des arcs $A(0)A(\frac{1}{2})$ et $A(\frac{1}{2})A(1)$ sont $m \cos^2 \alpha$ et $m \sin^2 \alpha$, α désignant l'angle $A(1)A(0)A(\frac{1}{2})$; leurs valeurs probables a priori, $A(1)$ étant inconnu, sont donc égales à $m/2$ (on remarque que, même si l'on adoptait pour α une loi de probabilité absolument quelconque, la somme de ces valeurs probables est toujours m). On en déduit, exactement comme dans le cas du mouvement brownien (§ 6, 2°), que l'ensemble des points communs aux deux arcs considérés a presque sûrement une mesure superficielle nulle; il en est de même a fortiori, quel que soit τ entre zéro et $\frac{1}{2}$, de l'ensemble des points communs aux arcs $A(\frac{1}{2} - \tau)A(\frac{1}{2})$ et $A(\frac{1}{2})A(\frac{1}{2} + \tau)$, qui n'est qu'une partie du précédent.

Nous prendrons $\tau = 1/64$ (pour la courbe Γ'_3 on pourrait prendre $1/16$). On voit aisément que, si $A(0)$, $A(\frac{1}{2})$ et $A(1)$ sont connus, les positions possibles pour chacun des points $A(\frac{1}{2} - \tau)$ et $A(\frac{1}{2} + \tau)$ recouvrent une aire entourant complètement le point $A(\frac{1}{2})$, et cela avec une densité de probabilité admettant au voisinage de ce point une borne inférieure positive.

Désignons par C_1 une quelconque des formes possibles de l'arc $A(\frac{1}{2})A(1)$. Les différents arcs $A(\frac{1}{2})A(\frac{1}{2} + \tau)$ possibles s'obtiennent en choisissant une position possible pour $A(\frac{1}{2} + \tau)$ et un arc C_1 ; $A(\frac{1}{2})A(\frac{1}{2} + \tau)$ sera l'arc semblable à C_1 allant de $A(\frac{1}{2})$ à $A(\frac{1}{2} + \tau)$. Supposant l'origine placée au point $A(\frac{1}{2})$, et représentant les points du plan par leurs affixes $X + iY$, nous désignerons par $U(t)$ ($\frac{1}{2} < t < 1$) l'affixe du point $A(t)$ de l'arc C_1 et par

$V(t)$ ($\frac{1}{2} < t < \frac{1}{2} + \tau$) celle du point $A(t)$ de l'arc $A(\frac{1}{2})A(\frac{1}{2} + \tau)$ semblable à C_1 et aboutissant en un point $A(\frac{1}{2} + \tau)$ d'affixe V_1 . On a évidemment

$$V = V(\frac{1}{2} + u\tau) = V_1 \frac{U[(1+u)/2]}{U(1)} \quad (0 < u < 1).$$

Supposons que C_1 ait une mesure superficielle positive; si V est donné et assez petit, et que l'on détermine V_1 par cette relation, les points d'affixes V_1 décrivent, quand u varie, une courbe transformée de C_1 ayant aussi une mesure superficielle positive, et de plus très voisine de $A(\frac{1}{2})$; il y aura donc une probabilité positive que $A(\frac{1}{2} + \tau)$ soit sur cette courbe, et que par suite l'arc $A(\frac{1}{2})A(\frac{1}{2} + \tau)$ contienne le point M d'affixe V ; c'est un point quelconque dans le voisinage de $A(\frac{1}{2})$.

Si alors il y avait une probabilité positive que Γ' ait une mesure superficielle positive, il en serait de même pour C_1 ; la circonstance que nous venons d'examiner se produisant avec une probabilité positive, la probabilité $\phi_1(M)$ que, $A(\frac{1}{2})$ étant connu, M appartienne à l'arc $A(\frac{1}{2})A(\frac{1}{2} + \tau)$, serait positive pour M assez voisin de $A(\frac{1}{2})$; il en serait de même de la probabilité $\phi_0(M)$ relative à l'arc $A(\frac{1}{2} - \tau)A(\frac{1}{2})$. Or, une fois $A(\frac{1}{2})$ choisi, ces arcs sont indépendants, et la probabilité que M appartienne à la fois à ces deux arcs aurait la valeur $\phi(M) = \phi_0(M)\phi_1(M)$ positive au voisinage de $A(\frac{1}{2})$. Son intégrale dans tout le plan, qui est la mesure superficielle probable de l'ensemble commun à ces deux arcs, quand $A(\frac{1}{2})$ est connu, serait positive. Cette conclusion étant vraie quel que soit le point $A(\frac{1}{2})$, sauf s'il occupait une des positions extrêmes $A(0)$ et $A(1)$, ce qui est infiniment peu probable, la valeur probable a priori de cet ensemble serait nulle, ce qui est en contradiction avec le résultat obtenu plus haut. Le résultat énoncé est donc établi.

Nous avons ainsi vérifié une fois de plus un fait évidemment très général: *quand l'oscillation brownienne n'est pas infinie, pour que la courbe étudiée recouvre une aire, il faut une organisation de ses détours infiniment petits que le hasard n'a aucune chance de produire. Le cas général est celui où la mesure superficielle de la courbe considérée est nulle.*

Nous avons, dans les trois cas étudiés dans ce travail (courbes C , Γ_3 et Γ') utilisé l'indépendance, au moins lorsque certains éléments aléatoires sont connus, d'un arc précédant ce point et d'un arc suivant ce point. Il est évident qu'une relation aussi précise que l'indépendance stochastique n'est pas nécessaire. Ainsi les probabilités $\phi_0(M)$ et $\phi_1(M)$ pourraient n'être pas indépendantes; si $\phi_1(M)$, quoique dépendant de l'arc $A(0)A(\frac{1}{2})$ supposé connu, avait une borne inférieure positive, la conclusion subsisterait.

Il y a donc lieu de penser que le principe général que nous venons d'indiquer peut s'appliquer à beaucoup d'autres schémas que ceux étudiés dans ce travail.

PARIS.

A GALOIS THEORY OF LINEAR SYSTEMS OVER COMMUTATIVE FIELDS.*¹

By REINHOLD BAER.

N. Jacobson² has recently succeeded in extending the Galois Theory from commutative fields to non-commutative fields. In accordance with the nowadays customary point of view he considers the Galois Theory as the theory of finite groups of automorphisms of commutative fields and of their fields of invariants. This theory contains the classical correspondence theorem of Galois as a simple special case. His fundamental condition which makes it possible to carry the commutative theory over to the non-commutative case is the restriction to finite groups of automorphisms without inner automorphisms $\neq 1$. His method consists in the application of the theory of simple rings without making much use of the commutative Galois Theory.

In this paper we give a different approach to Jacobson's theory. Our intent is to use the commutative Galois Theory ruthlessly and it turns out that beyond doing this one needs hardly more than the fact that a non-singular matrix with coefficients in a commutative field possesses an inverse; in particular we do not need any deeper facts concerning linear transformations or ideals.

Our method makes it possible to extend the theory in several directions. First we may investigate instead of non-commutative fields linear systems over commutative fields which need not have a finite basis over this field of reference and with one exception all the theorems of Galois Theory proper hold true in this framework. The exception which does not hold true may be easily derived in the case of fields from certain theorems concerning linear systems and is actually wrong for linear systems. Secondly we can prove that—at least for infinite linear systems—Jacobson's condition, properly phrased, is necessary for the validity of a Galois Theory. The rephrasing consists in substituting the concept "central-automorphism" for "inner automorphism"; and this is necessary, since only the former can be defined in the case of linear systems. This is the reason why we need the stronger hypothesis to obtain even those results which Jacobson is able to establish on the basis of his weaker condition. Thirdly we may prove quite general theorems which permit the

* Received August 4, 1939.

¹ Presented to the American Mathematical Society, September 1939.

² N. Jacobson, *Annals of Mathematics*, vol. 41 (1940), pp. 1-7.

transfer of the Galois Theory of any class of groups of automorphisms of commutative fields to linear systems whenever there exists a commutative Galois Theory so that in particular the Galois Theory of infinite algebraic extensions may be extended to linear systems; and these general "transfer" theorems are actually the starting point of our investigations.

Finally we give the elements of a theory of crossed products of non-commutative fields with finite groups of automorphisms. The standard theorems may easily be carried over. Only when proving a generalization of E. Noether's "Hauptgeschlechtssatz im minimalen" do we have to assume that the field in question be finite over its central so that we may use the theorem that central-automorphisms are inner automorphisms, a theorem that otherwise has no place in our theory. In this chapter as in the others Jacobson's work and ours overlap in many respects though the methods employed are rather different—his being strictly non-commutative, ours strictly commutative—and though neither obtains all the results of the other one.

CHAPTER I. Fundamentals and transfer.

1. If the set L is a commutative group with regard to an operation which is written as addition, if L contains elements different from 0, if F is a commutative³ field, and if there exists to every element f in F , x in L a uniquely determined element $fx = xf$ in L so that

- (a) $f(x + y) = fx + fy$ for f in F , x, y in L ,
- (b) $(f + g)x = fx + gx$ for f, g in F , x in L ,
- (c) $f(gx) = (fg)x$ for f, g in F , x in L ,
- (d) $1x = x$ for x in L and 1 the unit element in F ,

then L is called a *linear system over the field F* .

Note that (d) assures the absence of zero-divisors.

Dependence and independence of subsets of L with regard to F may be defined as usual. Every independent set is contained in a greatest independent set, every greatest independent set is a basis and any two greatest independent sets contain the same number of elements. These remarks clearly concern L as an abelian operator group with operators in F . However it is not this aspect of the matter that interests us primarily.

³ It may be noted that in some parts of Chapter I it is not necessary to assume that F be a commutative field, that the property of being a field will be sufficient for these considerations.

If S is any subset of L , then we denote by $Z(S)$ the set of all the elements z in F so that zs is in S for every s in S . The subset S of L is said to be *complete in L* , if

- (i) S is a subgroup of the addition group L_+ ,
- (ii) $S \neq 0$,
- (iii) $Z(S)$ is a subfield of F .

Most of the complete sets which we shall have to consider will satisfy an additional property:

- (iv) If f is an element in F so that there exists an element $s \neq 0$, satisfying: fs is an element in S , then f is an element in $Z(S)$.

If S is a subset of L , U a subset of F , then US is the subgroup of the additive group L_+ which is generated by all the elements us for u in U and s in S .

If S is complete in L , then the subset V of F is said to be *independent over S* , if $\sum_{i=1}^n v_i s_i = 0$ for v_i in V and s_i in S implies that all the s_i are 0.

(1.1) If S is complete in L , and if the subset V of F is independent over S , then V is independent over $Z(S)$.

Proof. Suppose that $\sum_{i=1}^n z_i v_i = 0$ for v_i in V and z_i in $Z(S)$. There exists in S an element $s \neq 0$. Hence all the elements $s_i = z_i s$ are in S . Thus we have $\sum_{i=1}^n s_i v_i = s \sum_{i=1}^n z_i v_i = 0$. Since V is independent over S , this implies that $0 = s_i = z_i s$. Hence $z_i = 0$, since $s \neq 0$.

The converse of (1.1) is in general not true. Hence we define:

The subset T of L is the direct product $U \times S$ of the subfield U of F by the subset S of L which is complete in L , if

- (1) $Z(S) \leq U$,
- (2) subsets of U are independent over $Z(S)$ if, and only if, they are independent over S ,
- (3) $T = US$.

It is a consequence of (1.1), that the conditions (2) and (3) may be condensed into the following condition:

- (0) A subset V of U is a basis of U over $Z(S)$ if, and only if, it is a basis

of T over S , i. e. every element t in T may be represented in one and only one way in the form:

$$t = \sum_{v \text{ in } V} s(v) v \text{ for } s(v) \text{ in } S,$$

where all the $s(v)$ —apart from a finite number of exceptions—are 0.

That T is generated in "adjoining" V to S is equivalent to (3); and the unicity of representation of elements in T is equivalent to (2).

A simple method for constructing subsets T of L so that L is the direct product of F and T is contained in the following statement.

THEOREM 1.2. *Suppose that L is a linear system over the commutative field F , and that the subset T of L is complete in L . Then L is the direct product of F and T if, and only if, every basis of the operator group T over $Z(T)$ is a basis of the operator group L over F .*

Proof. Suppose first that L is the direct product of F and T . Let B be some basis of the operator group T over $Z(T)$ so that T may be written: $T = \sum_{b \text{ in } B} Z(T)b$. Suppose furthermore that f_1, \dots, f_k are elements in F , b_1, \dots, b_k elements in B so that $\sum_{i=1}^k f_i b_i = 0$. Let A be some (linear) basis of F over $Z(T)$. Then there exist elements u_j in A , z_{ij} in $Z(T)$ so that $f_i = \sum_{j=1}^m z_{ij} u_j$ and thus we find that $0 = \sum_{i,j} z_{ij} u_j b_i = \sum_{j=1}^m t_j u_j$ where $t_j = \sum_{i=1}^k z_{ij} b_i$ is an element in T , since the b_i are in T and the z_{ij} are in $Z(T)$. Since L is the direct product of F and T , and since the u_j are elements in F which are independent over $Z(T)$, it follows that $0 = t_j = \sum_{i=1}^k z_{ij} b_i$. Since the elements b_i are part of a basis of the operator group T over $Z(T)$, they are independent over $Z(T)$ so that all the elements z_{ij} are 0, since they are in $Z(T)$. Thus all the f_i are 0, i. e. B is independent over F too. Since $L = FT$, every element in L depends on B (with coefficients in F) so that B is a basis of the operator group L over F .

Suppose now conversely that B is some basis of the operator group T over $Z(T)$ which is at the same time a basis of the operator group L over F . Then clearly $L = FT$. Suppose now that the elements f_1, \dots, f_k in F are (linearly) independent over $Z(T)$, and that the elements t_i in T satisfy: $0 = \sum_{i=1}^k t_i f_i$. Since B is a basis of T , there exist elements z_{ij} in $Z(T)$, b_j in B so that $t_i = \sum_{j=1}^m z_{ij} b_j$ and hence

$$0 = \sum_{i,j} z_{ij} b_j f_i = \sum_{j=1}^m \left[\sum_{i=1}^k z_{ij} f_i \right] b_j.$$

Since the elements b_j are part of a basis of L over F , it follows that $0 = \sum_{i=1}^k z_{ij} f_i$, as these are elements in F ; and since the f_i are independent elements in F over $Z(T)$, it follows that the elements z_{ij} in $Z(T)$ are 0. Thus all the t_i are 0; and this implies that every set in F which is independent over $Z(T)$ is at the same time independent over T . Hence L is the direct product of F and T ; and this completes the proof. As a matter of fact we have proved slightly more namely the

COROLLARY 1.3. *If L is a linear system over the commutative field F , and if the subset T of L is complete in L , then the following propositions are equivalent.*

(A)
$$L = F \times T.$$

(B) *There exists at least one basis of F over $Z(T)$ which is a basis of L over T .*

(C) *Every basis of the operator group T over $Z(T)$ is a basis of the operator group L over F .*

(D) *There exists at least one basis of the operator group T over $Z(T)$ which is a basis of the operator group L over F .*

2. In this section we shall introduce the concept of automorphism of a linear system which, of course, will differ from the concept of automorphism of an operator group.

(2.1) *If L is a linear system over the (commutative) field F , then there exists to a given automorphism g of the additive group L , at most one automorphism h of the field F so that*

$$(fx)^g = f^h x^g \text{ for } f \text{ in } F \text{ and } x \text{ in } L.$$

Proof. Suppose that h and k are two automorphisms of the field F so that $f^h x^g = (fx)^g = f^k x^g$ for f in F and x in L . There exists in L an element $w \neq 0$; and $w^g \neq 0$, since g is an automorphism of L . Hence $f^h w^g = f^k w^g$ or $(f^h - f^k) w^g = 0$ and this implies $f^h = f^k$ for every f in F or $h = k$.

Consequently we define: *The transformation g of L is an automorphism⁴ of the linear system L over the field F , if*

⁴These automorphisms of linear systems over fields are often termed semi-linear transformations; cp. e.g. N. Jacobson, *Annals of Mathematics*, vol. 38 (1937), 484-507.

(*) g is an automorphism of the additive group L , and if

(**) there exists one (and only one) automorphism h of the field F so that $(fx)^g = f^h x^g$ for f in F and x in L .

If g is an automorphism of the linear system L over the field F , then we say of the uniquely determined automorphism h of the field F which occurs in (**) that it is induced by g and put $h = g'$. If G is some set of automorphisms of L , then we denote by G' the set of all the g' for g in G .

If u, v are both automorphisms of L , then $(uv)' = u'v'$. If G is a group of automorphisms of L , then a homomorphism of G upon the group G' is defined in mapping the element g in G upon the element g' in G' ; and this homomorphism of G upon G' is an isomorphism between G and G' if, and only if, $g' = 1$ implies $g = 1$ (for elements g in G).

An automorphism which leaves all the elements in some set Y invariant is called a Y -automorphism. If the subset S of the linear system L over the field F is complete in L , and if g is an S -automorphism of L , then g' is a $Z(S)$ -automorphism of L .

If G is a group of automorphisms of L , then (L, G) consists of all the elements in L which are left invariant by every automorphism in G ; and (F, H) is defined accordingly.

If S is a subset of L , then $(S < L)$ is the group of all the S -automorphisms of L ; and $(T < F)$ is defined accordingly. These operations satisfy (as usual):

$$G \leq ((L, G) < L) \quad \text{and} \quad S \leq (L, (S < L)).$$

Since furthermore $G \leq H$ implies $(L, H) \leq (L, G)$, and since $S \leq T$ implies $(T < L) \leq (S < L)$, it follows that

$$(S < L) = ((L, (S < L)) < L) \quad \text{and} \quad (L, G) = (L, ((L, G) < L)).$$

(2.2) Let G be a group of automorphisms of the linear system L over the commutative field F .

(a) If every (F, G') -automorphism of F is induced by at most one (L, G) -automorphism of L , then $(L, G) \neq 0$.

(b) If $(L, G) \neq 0$, then

(b.1) $Z((L, G)) = (F, G')$,

(b.2) (L, G) is complete in L ,

(b.3) the element f in F belongs to $Z((L, G))$ whenever there exists an element $t \neq 0$ in (L, G) so that ft is in (L, G) .

Proof. Assume that every (F, \mathbf{G}') -automorphism of F is induced by at most one (L, \mathbf{G}) -automorphism of L and that $(L, \mathbf{G}) = 0$. If f is an element $\neq 0$ in F , then an automorphism of L is defined by $x^g = fx$ for x in L , since $(rx)^g = frx = rx^g$ for r in F . Hence g is an (L, \mathbf{G}) -automorphism of L satisfying $g' = 1$, and this implies $g = 1$ so that $f = 1$, i. e. F consists of 0 and 1 only. Consequently every $g' = 1$ so that every $g = 1$; and hence $(L, \mathbf{G}) = L \neq 0$ which is a contradiction.

Assume now that $(L, \mathbf{G}) \neq 0$, and that f is an element in F , $t \neq 0$ an element in $(L, \mathbf{G}) = T$ and that ft is in T too. If g is any automorphism in G , then $ft = (ft)^g = f^g t$ so that $f = f^g$ for every g' in \mathbf{G}' . Hence f is an element in (F, \mathbf{G}') and in particular $Z(T) \leq (F, \mathbf{G}')$.—If z is an element in (F, \mathbf{G}') , t any element in T , then $(zt)^g = zt$ for every g in \mathbf{G} so that zt is in T and consequently z is in $Z(T)$. Hence $Z(T) = (F, \mathbf{G}')$ so that $Z(T)$ is a field and T is complete in L .

THEOREM 2.3. *Suppose that L is a linear system over the commutative field F , and that the subset T of L is complete in L .*

- (a) *If L is the direct product of F and T , then every $Z(T)$ -automorphism of F is induced by one and only one T -automorphism of L .*
- (b) *$L = FT$ if, and only if, the identity is the only T -automorphism of L which induces the identity in F .*
- (c) *If $Z(T) = (F, (T < L)')$, then every independent subset of the operator group T over $Z(T)$ is independent over F too.*

Proof. If L is the direct product of F and T , and if B is a basis of F over $Z(T)$, then B is a basis of L over T . If x is any element in L , then there exist therefore uniquely determined elements $t(x, b)$ in T so that only a finite number of $t(x, b)$ are different from 0, and so that $x = \sum_{b \text{ in } B} t(x, b)b$. If g and h are two T -automorphisms of L so that $g' = h'$, then $x^g = \sum_{b \text{ in } B} t(x, b)b^{g'}$ $= \sum_{b \text{ in } B} t(x, b)b^{h'} = x^h$ so that $g = h$. If conversely u is a $Z(T)$ -automorphism of F , then a T -automorphism v of L is defined by $x^v = \sum_{b \text{ in } B} t(x, b)b^u$ and clearly $v' = u$. This proves (a).

$L^* = FT$ is in any case an admissible subgroup of the operator group L over F . Thus every basis of the operator group L^* over F is contained in some basis B of L over F . If $L^* < L$, then there exists in B an element w which is not contained in L^* . As $T \neq 0$, there exists in T an element $t \neq 0$; and there exists one and only one automorphism g so that $g' = 1$, $b = b^g$ for $b \neq w$ in B , $w^g = w + t$. Since g is a T -automorphism, and since $g \neq 1$,

it follows that $FT \neq L$ implies the existence of a T -automorphism $\neq 1$ which induces the identity in F ; and this proves (b), since g is the identity on FT , if g is a T -automorphism such that $g' = 1$.

Suppose now that $Z(T) = (F, (T < L)')$ and that S is a subset of T which is independent over $Z(T)$. If S would not be independent over F , then S would contain a finite subset which is dependent over F , and amongst these there would be a smallest one, say s_1, \dots, s_k . There exist therefore elements f_i , not all of them 0, so that $0 = \sum_{i=1}^k s_i f_i$ and since the s_i form a smallest dependent subset of S , none of the f_i is 0. If g is any T -automorphism of L , then $0 = \sum_{i=1}^k s_i f_i g'$ and consequently $0 = \sum_{i=2}^k s_i (f_i f_1 g' - f_i g' f_1)$; and this implies $f_i f_1 g' = f_i g' f_1$, since the s_i form a smallest [over F] dependent subset of S . Hence $f_i f_1^{-1}$ is invariant under all the g' in $(T < L)'$ and belongs therefore to $Z(T)$. Since $f_1 \neq 0$, we find therefore that $0 = \sum_{i=1}^k s_i (f_i f_1^{-1})$ where all the coefficients are elements $\neq 0$ in $Z(T)$. Hence the s_i and therefore S would be dependent over $Z(T)$ and this is impossible so that (c) holds true too.

3. The problem of a Galois Theory of linear systems will be reduced by means of the theorems in this section to the corresponding problems of Galois Theory in commutative fields.

THEOREM 3.1. *The subset T of the linear system L over the commutative field F satisfies*

- (a) $T = (L, (T < L))$ and
- (b) 1 is the only automorphism g in $(T < L)$ so that $g' = 1$,
if, and only if,
 - (i) T is complete in L ,
 - (ii) $Z(T) = (F, (Z(T) < F))$,
 - (iii) L is the direct product of F and T .

Proof. Suppose first that the conditions (a) and (b) are satisfied by T . Then it follows from (2.2) that T is complete in L (condition (i)!) and that $Z(T) = (F, (T < L)')$. Since therefore $(T < L)' \leq (Z(T) < F)$, it follows that

$$Z(T) \leq (F, (Z(T) < F)) \leq (F, (T < L)') = Z(T)$$

and this proves (ii).

Suppose now that B is a basis of the operator group T over the field $Z(T)$. Then it follows from Theorem 2.3, (c) that B is independent over F so that B is a basis of the operator group L over F , since it follows from (b) and Theorem 2.3, (b) that $L = FT$. Hence L is the direct product of F and T by Theorem 1.2.

Suppose now conversely that the conditions (i) to (iii) are satisfied by T . Then it is a consequence of Theorem 2.3, (a) that (b) holds true, and that $(Z(T) < F) = (T < L)'$. Let now B be a basis of the operator group T over $Z(T)$. Then B is by Theorem 1.2 and condition (iii) a basis of the operator group L over F . If w is any element in $(L, (T < L))$, then there exist elements f_i in F and elements b_i in B so that $w = \sum_{i=1}^k f_i b_i$. If v is any automorphism in $(Z(T) < F)$, then there exists a T -automorphism g of L so that $g' = v$; and hence we find that $\sum_{i=1}^k f_i v b_i = w^g = w = \sum_{i=1}^k f_i b_i$ so that $f_i = f_i v$ for every v in $(Z(T) < F)$. All the elements f_i are therefore in $(F, (Z(T) < F))$; and since this field is equal to $Z(T)$ by condition (ii), it follows that all the f_i are in $Z(T)$ and that w is in T , since all the b_i are in T . Hence $T \leq (L, (T < L)) \leq T$, i. e. (a) holds true too.

THEOREM 3.2. *The group G of automorphisms of the linear system L over the commutative field F satisfies*

(a) $G = ((L, G) < L)$ and

(b) 1 is the only automorphism g in G so that $g' = 1$

if, and only if,

(i) $G' = ((F, G') < F)$ and

(ii) every (F, G') -automorphism of F is induced by one and only one (L, G) -automorphism of L .

Proof. Suppose first that (a) and (b) are satisfied by G . Put $T = (L, G)$ so that $G = (T < L)$ and $T = (L, (T < L))$ by (a). Hence it follows from (b) and Theorem 3.1 that T is complete in L , that $Z(T) = (F, (Z(T) < F))$ and that L is the direct product of F and T . Now it is a consequence of Theorem 2.3 that every $Z(T)$ -automorphism of F is induced by one and only one T -automorphism of L and hence (ii) holds true, since

$$Z(T) = (F, (Z(T) < F)) = (F, (T < L)')$$

by (2.2). Finally it follows now from (a) that

$$G' = ((L, G) < L)' = (T < L)' = ((F, G') < F)$$

so that (i) holds true too.

Assume conversely that \mathbf{G} satisfies (i) and (ii). If the automorphism w is in $((L, \mathbf{G}) < L)$, then w' is in $(Z((L, \mathbf{G})) < F)$; and since $Z((L, \mathbf{G})) = (F, \mathbf{G}')$ by (ii) and (2.2), w' is in $((F, \mathbf{G}') < F)$ which group equals \mathbf{G}' by (i). Hence it follows from (ii) that w is in \mathbf{G} so that (a) and (b) are consequences of (i) and (ii).

COROLLARY 3.3. *Suppose that the group \mathbf{G} of automorphisms of the linear system L over the commutative field F satisfies the conditions (a) and (b) of Theorem 3.2, and that \mathbf{H} is a subgroup of \mathbf{G} . Then $\mathbf{H} = ((L, \mathbf{H}) < L)$ if, and only if, $\mathbf{H}' = ((F, \mathbf{H}') < F)$.*

This follows from Theorem 3.2, since \mathbf{H} satisfies condition (b) of Theorem 3.2 as a subgroup of \mathbf{G} , and since $(F, \mathbf{G}') \leq (F, \mathbf{H}')$ implies that \mathbf{H} satisfies condition (ii) of Theorem 3.2 as \mathbf{G} satisfies this condition.

THEOREM 3.4. *Suppose that L is a linear system over the field F , that T is a subset of L , that $T = (L, (T < L))$, that $\mathbf{1}$ is the only T -automorphism g of L with $g' = \mathbf{1}$, and that B is a set between T and L .*

(A) $B = (L, (B < L))$ if, and only if, there exists a field R between $Z(T)$ and F so that $R = (F, (R < F))$ and so that $B = RT$.

(B) If R is a field between $Z(T)$ and F so that $R = (F, (R < F))$, then $R = Z(RT)$ and RT is the direct product of R and L .

Proof. Suppose first that there exists a field R between $Z(T)$ and F so that $R = (F, (R < F))$ and so that $B = RT$. Then $R \leq Z(B)$. Let z be an element in $Z(B)$, t an element $\neq 0$ in T and v an R -automorphism of F . Then there exists by Theorem 3.1 and 2.3 a T -automorphism g of L so that $g' = v$. Since g leaves all the elements in T invariant, and since v leaves all the elements in R invariant, g leaves all the elements in RT invariant. Since $T \leq RT$, and since z is in $Z(RT)$, zt is an element in RT so that $zt = (zt)^g = z^v t$ and $z = z^v$ since $t \neq 0$. Hence z is in $(F, (R < F)) = R$ so that $R = Z(TR)$. Hence (B) holds true, since subsets of $R \leq F$ that are independent over $Z(T)$ are independent over T .

Let now S be a basis of the operator group T over $Z(T)$. It follows from Theorem 3.1 and Theorem 1.2 that S is a basis of the operator group L over F . If w is an element in $(L, (RT < L))$, then there exist elements f_i in F and elements s_i in S so that $w = \sum_{i=1}^k f_i s_i$. If v is any R -automorphism of F , then there exists again a T -automorphism g of L so that $g' = v$; and g is an RT -automorphism of L . Hence $\sum_{i=1}^k f_i s_i = w = w^g = \sum_{i=1}^k f_i^v s_i$ so that $f_i = f_i^v$,

since the s_i are in T and are independent over F . The elements f_i are therefore in $(F, (R < F)) = R$ so that w is in RT , i. e. $RT = (L, (RT < L))$.

Assume now conversely that the set B between T and L satisfies $B = (L, (B < L))$. There exists no B -automorphism of L , inducing 1 in F , except 1 , since there exists no T -automorphism $\neq 1$ of L which induces the identity in F . Hence it follows from Theorem 3.1 that B is complete in L , that $Z(B) = (F, (Z(B) < F))$ and that L is the direct product of F and B . Consequently $B^* = Z(B)T \leq B$; and it follows from what has been proved in the first two paragraphs that $Z(B^*) = Z(B)$ and that $B^* = (L, (B^* < L))$. It is a consequence of Theorem 3.1 and Theorem 2.3 that every $Z(T)$ -automorphism of F is induced by one and only one T -automorphism of L so that $(B < L)' = (Z(B) < F) = (Z(B^*) < F) = (B^* < L)'$ and therefore $(B < L) = (B^* < L)$ and finally $B^* = (L, (B^* < L)) = (L, (B < L)) = B$ and this completes the proof of (A).

THEOREM 3.5. *Suppose that T is a subset of the linear system L over the field F , that $T = (L, (T < L))$, that the identity is the only T -automorphism of L which induces the identity in F , and that the set B between T and L is complete in L . Then B satisfies:*

- (a) $B = B^*$ for every T -automorphism g of L ,
- (b) every T -automorphism of the linear system B over the field $Z(B)$ is induced by some automorphism of L ,
- (c) $(F, (B < L)') = Z(B)$

if, and only if, the following conditions are satisfied by B :

- (i) $(B < L)$ is a normal subgroup of $(T < L)$,
- (ii) every $Z(T)$ -automorphism of $Z(B)$ is induced by automorphisms of F ,
- (iii) $B = (L, (B < L))$.

Proof. We note first that $(B^* < L) = g^{-1}(B < L)g$ for every automorphism g of L . This shows that (i) is a consequence of (a). If conversely (i) and (iii) are satisfied, then $B^* \leq (L, (B^* < L)) = (L, (B < L)) = B$ for every T -automorphism g of L . This implies that $B \leq B^{g^{-1}}$ for every T -automorphism g of L and therefore we have $B \Rightarrow B^*$ for every T -automorphism g of L , i. e.

(a) is a consequence of (i) and (iii).

If (iii) is true, then it follows from Theorem 3.4 that $Z(B) = (F, (Z(B) < F))$ and that $B = Z(B) \times T$. It follows from Theorem 2.3 that every $Z(T)$ -automorphism of $Z(B)$ is induced by one and only one T -automorphism of B . If now g is any T -automorphism of B , then g' is a $Z(T)$ -automorphism of $Z(B)$. If (ii) holds true, then there exists an automorphism u of F which induces g' in $Z(B)$. There exists by Theorem 2.3 one and only one T -automorphism h of L so that $h' = u$. It is a consequence of (a) that h induces an automorphism k in B . Since clearly $k' = g'$, it follows that $k = g$, as every $Z(T)$ -automorphism of $Z(B)$ is induced by one and only one T -automorphism of B . Thus (b) is a consequence of (i) to (iii).

Suppose now that u is a $Z(B)$ -automorphism of F . As u is a $Z(T)$ -automorphism of F , there exists one and only one T -automorphism v of L so that $v' = u$. Since $B = Z(B) \times T$, it follows that v is a B -automorphism of L , and this shows that $(Z(B) < F) = (B < L)'$. Since we already proved that $Z(B) = (F, (Z(B) < F))$, condition (c) is a consequence of (i) to (iii).

We assume now that conditions (a) to (c) are satisfied by B . Let S be any basis of the operator group T over $Z(T)$. Then S is a basis of the operator group L over F so that S is independent over $Z(B)$. Hence S is contained in a basis S^* of the operator group B over $Z(B)$. But it follows from (c) and Theorem 2.3, (c) that S^* is independent over F too. Consequently $S = S^*$ and B is the direct product of $Z(B)$ and T , as follows from Theorem 1.2.

Since $(B < L)' \leq (Z(B) < F)$, and since therefore

$$Z(B) \leq (F, (Z(B) < F)) \leq (F, (B < L)') = Z(B)$$

by (c) or $Z(B) = (F, (Z(B) < F))$, it follows now from Theorem 3.4 that $B = (L, (B < L))$.

Suppose finally that u is a $Z(T)$ -automorphism of $Z(B)$. Since B is the direct product of $Z(B)$ and T , there exists by Theorem 2.3 one and only one T -automorphism v of B so that $v' = u$. It is a consequence of (b) that there exists an automorphism g of L which induces v in B . Then the automorphism g' of F induces $v' = u$ in $Z(B)$. This completes the proof of the fact that (i) to (iii) are consequences of (a) to (c).

CHAPTER II. Galois Theory.

4. In this section we state the *finite, commutative Galois Theory* in the form most convenient for our purposes.⁵

(4.1) Suppose that K is a subfield of the (commutative) field F . Then there exists a finite group H of automorphisms of the field F so that

$$K = (F, H)$$

if, and only if, F is finite, normal and separable over K .

(4.2) If H is a finite group of automorphisms of the (commutative) field F , then

$$H = ((F, H) < F).$$

(4.3) If F is finite, normal and separable over its subfield K , then

$$B = (F, (B < F))$$

for every field B between K and F , i. e. F is finite, normal and separable over every field B between K and F .

(4.4) If F is finite, normal and separable over its subfield K , and if B is a field between K and F , then a necessary and sufficient condition for B to be finite, normal and separable over K is that

$$(B < F) \text{ is a normal subgroup of } (K < F),$$

and then $(K < B)$ and $(K < F)/(B < F)$ are essentially the same.

(4.5) If F is finite, normal and separable over its subfield K , then there exists in F an element b so that the elements b^h for h in $(K < F)$ form a basis of F over K . (Existence of a normal basis).⁶

It may finally be mentioned that finite and separable extensions are

⁵ Apart from the text-books on modern algebra one should consult the following papers in which the theory has been presented in a form similar to the one sketched here. R. Baer, *Mathematische Zeitschrift*, vol. 33 (1931), pp. 451-479; R. Baer, *American Journal of Mathematics*, vol. 59 (1937), pp. 869-888; W. Krull, *Mathematische Annalen*, vol. 100 (1929), pp. 687-698; E. Steinitz, *Algebraische Theorie der Körper*. Neu herausgegeben und mit einem Anhang: *Abriss der Galoisschen Theorie* versehen von Reinhold Baer und Helmut Hasse. Berlin, 1930.

⁶ A complete proof of this theorem has first been given by M. Deuring, *Mathematische Annalen*. All the proofs published so far use extensively the theory of representations. There exists however an unpublished proof by E. Artin which uses but elementary means from the theory of fields so that this theorem may now be considered a part of Galois Theory proper.

simple extensions, and that the degree of a finite, normal and separable extension is exactly the order of its group, and that the matrix (b_i^g) possesses an inverse, if the b_i form a basis, the g the group of a finite, normal and separable extension; finally that every automorphism of a subfield of a normal extension is induced by an automorphism of the extension field.

5. In this section some remarks, concerning *matrices and linear equations*, shall be given which will prove useful in the future.

Let L be a linear system over the field F . If B is a matrix of n rows and n columns with coefficients in F , and if X is a matrix of n rows and one column with coefficients in L , then $BX (= (b_{ik})(x_j)) = (\sum_{k=1}^n b_{ik}x_k)$ is a matrix of n rows and one column with coefficients in L .

If A and B are two matrices of n rows and n columns, both with coefficients in F , and if X is a matrix of n rows and one column with coefficients in L , then one verifies readily that

$$A(BX) = (AB)X.$$

If in particular E is the unit-matrix in F , then $EX = X$.

It is now possible to write the system of linear equations

$$(+)\quad \sum_{k=1}^n b_{ik}x_k = c_i \text{ for } i=1, \dots, n, \ b_{ik} \text{ in } F, \ c_i \text{ in } L,$$

in the matrix form: $(b_{ik})(x_k) = (c_i)$. The solutions x_k of $(+)$ should be looked for in L . If in particular the matrix $(b_{ik}) = B$ is non-singular, i.e. if the determinant of B is different from 0, then there exists the inverse matrix B^{-1} to B ; and the system $(+)$ of linear equations has *one and only one system of solutions* x_k in L , since $(b_{ik})(x_k) = (c_i)$ if, and only if,

$$B^{-1}(c_i) = B^{-1}(b_{ik})(x_k) = E(x_k) = (x_k).$$

6. Since the Galois Theory of finite groups of automorphisms is fully developed, it is possible to derive stronger theorems in the case of finite groups of automorphisms than the theorems of section 3.

THEOREM 6.1. *Let T be a subset of the linear system L over the commutative field F . Then there exists a finite group G of automorphisms of L so that*

(1) *the identity is the only automorphism in G which induces the identity in F ,*

$$(2) \quad T = (L, G)$$

if, and only if,

- (i) T is complete in L ,
- (ii) L is the direct product of F and T ,
- (iii) F is finite, normal and separable over $Z(T)$.

Proof. Assume first that the finite group G of automorphisms of L satisfies the conditions (1) and (2), and that $T = (L, G)$. Then G and G' are isomorphic finite groups. Hence F is finite, normal and separable over (F, G') by (4.1).

Now let b_1, \dots, b_n be a basis of F over (F, G') . Then G and G' both contain n elements; and there exists an inverse M to the matrix $(b_i g')$ where the row-index g' runs over all the elements in G' . If u is any element in L , then the system

$$(+)\quad \sum_{i=1}^n b_i g' x_i = u^g \text{ for } g \text{ in } G$$

of linear equations possesses one and only one system of solutions x_i in L , namely—in matrix-notation— $(x_i) = M(u^g)$. If h is any automorphism in G , then (x_i^h) satisfies

$$\sum_{i=1}^n b_i g' h' x_i^h = u^{gh} \text{ for } g \text{ in } G.$$

Since gh runs over all the elements in the group G when g takes all the values in G , it follows that the x_i^h are another solution of (+); and since (+) possesses but one solution it follows that $x_i = x_i^h$ for every h in G so that the x_i are actually elements in $T = (L, G)$. This implies in particular that $L = FT$.—If one applies this result concerning (+) on $u = 0$, then it follows that the b_i are independent over T , since $\sum_{i=1}^n b_i t_i = 0$ with t_i in T implies that the equations $\sum_{i=1}^n b_i g' t_i = 0$ for g in G are satisfied, and since the only solutions of these equations are $t_i = 0$.

Since $L = FT$, $T \neq 0$; and it follows from (2.2) that $Z(T) = (F, G')$, that T is complete in L , and therefore from the results of the first paragraph of the proof that L is the direct product of F and T . The conditions (i) to (iii) are therefore satisfied by T .

Assume conversely that the conditions (i) to (iii) are satisfied by T . Then it follows from (4.3) that $Z(T) = (F, (Z(T) < F))$ and it follows therefore from Theorem 3.1 that $(T < L)$ satisfies condition (1) and that $T = (L, (T < L))$. Since $(T < L)$ satisfies (1), $(T < L)$ and $(T < L)'$ are isomorphic groups, so that $(T < L)$ is finite, since F is finite over $Z(T)$, and since $(T < L)'$ is a subgroup of $(Z(T) < F)$. Thus the existence of a

finite group \mathbf{G} of automorphisms of L , satisfying (1) and (2), is a consequence of (i) to (iii).

An alternative proof for this last inference may be given, as this second proof does not use Theorem 3.1. If (i) to (iii) are satisfied by T , then it follows from (4.3) that $Z(T) = (F, (Z(T) < F))$. Let b be an element in F so that the elements b^g for g in $(Z(T) < F)$ form a basis of F over $Z(T)$ (cp. (4.5)!). The elements b^g form a basis of L over T —by (ii)—and it follows from Theorem 2.3 that $(T < L)$ satisfies (1) and that $(Z(T) < F) = (T < L)'$ so that $(T < L) = \mathbf{G}$ is finite. If finally x is an element in $(L, (T < L))$, then there exist elements $t(g)$ in T so that $x = \sum_{g \in \mathbf{G}} t(g)b^g$. Consequently $x = \sum_{g \in \mathbf{G}} t(g)b^{g'h'}$ for every h' in \mathbf{G} ; and this implies that all the $t(g)$ are equal to a fixed element t in T so that $x = t \sum_{g \in \mathbf{G}} b^g = tz$ for z in $Z(T)$. Hence t is in T and consequently $T = (L, (T < L))$.

COROLLARY 6.2. Suppose that the subset T of the linear system L over the commutative field F is complete in L , and that F is finite, normal and separable over $Z(T)$.

(a) If L is the direct product of F and T , then $(T < L)$ is a finite group of automorphisms of L , the identity is the only T -automorphism of L which induces the identity in F and $T = (L, (T < L))$.

(b) L is the direct product of F and T if, and only if, every $Z(T)$ -automorphism of F is induced by one and only one T -automorphism of L .

Proof. (a) has already been verified in the proof of Theorem 6.1.—That the condition of (b) is necessary, follows from Theorem 2.3. If on the other hand every $Z(T)$ -automorphism of F is induced by one and only one T -automorphism of F , then $(T < L)' = (Z(T) < F)$ and therefore $Z(T) = (F, (Z(T) < F)) = (F, (T < L)')$ by (4.3). Hence it follows from Theorem 2.3, (b), (c) that L is the direct product of F and T .

THEOREM 6.3. If the identity is the only automorphism in the finite group \mathbf{G} of automorphisms of the linear system L over the commutative field F which induces the identity in F , then $\mathbf{G} = ((L, \mathbf{G}) < L)$.

Proof. It is a consequence of Theorem 6.1 that (L, \mathbf{G}) is complete in L , that F is finite, normal and separable over $Z((L, \mathbf{G}))$ and that L is the direct product of F and (L, \mathbf{G}) . Hence it follows from Corollary 6.2, (b) that every automorphism in $(Z((L, \mathbf{G})) < F)$ is induced by an (L, \mathbf{G}) -automorphism of L so that $((L, \mathbf{G}) < L)' = (Z((L, \mathbf{G})) < F)$, and it is a

consequence of (2.2) that $Z((L, \mathbf{G})) = (F, \mathbf{G}')$. Now it follows from (4.2) that $\mathbf{G}' = ((F, \mathbf{G}') < F) = (Z((L, \mathbf{G})) < F) = ((L, \mathbf{G}) < L)'$. Since \mathbf{G}' is finite, and since every $Z((L, \mathbf{G}))$ -automorphism of F is induced by one and only one (L, \mathbf{G}) -automorphism of L , this implies that $\mathbf{G} = ((L, \mathbf{G}) < L)$.

THEOREM 6.4. *Suppose that L is a linear system over the commutative field F , that the subset T of L is complete in L , that L is the direct product of F and T , that F is finite, normal and separable over $Z(T)$, and that B is a set between T and L .*

(A) $B = (L, (B < L))$ if, and only if, there exists a field R between $Z(T)$ and F so that $B = RT$.

(B) If R is a field between $Z(T)$ and F , then $R = Z(RT)$.

This is a consequence of Corollary 6.2 and of Theorem 3.4, since every field R between $Z(T)$ and F satisfies $R = (F, (R < F))$ by (4.3).

THEOREM 6.5. *Suppose that L , F , T and B satisfy the hypotheses of Theorem 6.4, and that B is complete in L . Then B satisfies*

(a) $(B < L)$ is a normal subgroup of $(T < L)$, and

(b) $B = (L, (B < L))$

if, and only if,

(i) $B = B^g$ for every T -automorphism g of L , and

(ii) $(F, (B < L)') = Z(B)$.

Proof. Every $Z(T)$ -automorphism of $Z(B)$ is induced by some automorphism of F , since $Z(B)$ is between $Z(T)$ and F , and since F is finite and normal over $Z(T)$. Thus the above conditions (a) and (b) imply the conditions (i) to (iii) of Theorem 3.5 and consequently the above conditions (i) and (ii).—If conversely the above conditions (i) and (ii) are satisfied, then (a) is a consequence of $(B^g < L) = g^{-1}(B < L)g$. Since $(T < L)$ is finite, $(B < L)$ and $(B < L)'$ are both finite, and hence it follows from (ii) and (4.2) that

$$(B < L)' \leq (Z(B) < F) = ((F, (B < L)') < F) = (B < L)'$$

or

$$(B < L)' = (Z(B) < F).$$

By (i) every T -automorphism g of L induces a T -automorphism g^* in the linear system B over $Z(B)$. If $g^{*'} = 1$, then g' is in $(Z(B) < F)$ and therefore in $(B < L)'$ so that g is in $(B < L)$, i. e. $g^* = 1$. The group \mathbf{G}^* of

these automorphisms \mathbf{g}^* is finite, satisfies condition (1) and (2) of Theorem 6.1, since $(B, \mathbf{G}^*) = (L, (T < L)) = T$. Hence it follows from Theorem 6.1 that $B = Z(B)T$, and it follows from Theorem 6.4 that $B = (L, (B < L))$.

The following theorem is some sort of a converse to Theorem 6.3.

THEOREM 6.6. *Suppose that the linear system L over the commutative field F contains an infinity of elements, that \mathbf{G} is a finite group of automorphisms of L and that $(L, \mathbf{G}) \neq 0$. Then $\mathbf{G} = ((L, \mathbf{G}) < L)$ if, and only if, the identity is the only automorphism in \mathbf{G} which induces the identity in F .*

Proof. The sufficiency of the condition is a consequence of Theorem 6.3.—Suppose now that the identity is not the only automorphism in \mathbf{G} which induces the identity in F . Then denote by \mathbf{W} the set of all the elements w in \mathbf{G} so that $w' = 1$. Clearly \mathbf{W} is a normal subgroup of \mathbf{G} . Let $V = (L, \mathbf{W})$. Then $Z(V) = F$ and the automorphisms in \mathbf{G} induce in V a finite group \mathbf{G}^* of automorphisms of the linear system V over F . This group \mathbf{G}^* is essentially the same as \mathbf{G}/\mathbf{W} . Since by the construction of V the identity is the only automorphism in \mathbf{G}^* which induces the identity in V , it follows from Theorem 6.3 that $\mathbf{G}^* = ((V, \mathbf{G}^*) < V)$ and it may be noted that $(V, \mathbf{G}^*) = (L, \mathbf{G}) = T$, $V = FT$.

Since $\mathbf{W} \neq 1$, $V < L$. Since V is an admissible subgroup of the operator group L over F , there exists a basis B of L over F which contains a basis U of V over F . Clearly $U < B$. Now we distinguish two cases.

Case 1. V contains an infinity of elements. Let d be some element in B that is not contained in U . If v is an element $\neq 0$ in V , then an automorphism $\mathbf{g} = \mathbf{g}(v)$ is defined by the conditions: $\mathbf{g}' = 1$, $d^{\mathbf{g}} = d + v$, $b = b^{\mathbf{g}}$ for $b \neq d$ in B . Each $\mathbf{g}(v)$ is a V -automorphism and therefore a T -automorphism of L . Since there exists an infinity of automorphisms $\mathbf{g}(v)$, it follows that $(T < L)$ is infinite so that the finite group \mathbf{G} is certainly smaller than $((L, \mathbf{G}) < L)$.

Case 2. V contains but a finite number of elements. Then there exists in B an infinity of elements d which are not contained in V and therefore not in U , since it follows from the finiteness of V and from $V \neq 0$, that the field F contains but a finite number of elements. Let v be some element $\neq 0$ in V . If d is an element in B that is not contained in U , then an automorphism $\mathbf{h} = \mathbf{h}(d)$ is defined by the conditions: $\mathbf{h}' = 1$, $d^{\mathbf{h}} = d + v$, $b = b^{\mathbf{h}}$ for $b \neq d$ in B . All these automorphisms are different. They are V -automorphisms and therefore they are T -automorphisms of L . Consequently $(T < L)$ is infinite and therefore different from the finite group \mathbf{G} . Hence we have proved that

$\mathbf{G} = ((L, \mathbf{G}) < L)$ implies that the identity is the only automorphism in \mathbf{G} that induces the identity in F , provided L is infinite.

Remark. If L is a finite system, then every group of automorphisms of L is finite. In this case—using the notations of the proof of the preceding theorem— $\mathbf{G} = ((L, \mathbf{G}) < L)$ if, and only if, $\mathbf{W} = ((L, \mathbf{W}) < L)$.

7. In this section a short discussion is given of possibilities of extending Theorem 6.4 and Theorem 6.5. The following assumptions will be made throughout this section. L is a linear system over the commutative field F ; \mathbf{G} is a finite group of automorphisms of L so that the automorphism g in \mathbf{G} satisfies $g' = 1$ if, and only if, $g = 1$; $T = (L, \mathbf{G})$. Then we prove:

There exists a set W between T and L so that

$$(a) \quad T < W \leq L,$$

$$(b) \quad Z(T) = Z(W),$$

(c) *W is complete in L and $Z(W)$ contains every element f in F to which there exists an element $w \neq 0$ in W so that fw is in W*

if, and only if, the order of \mathbf{G} is greater than 2 and the rank of the operator group L over F is greater than 1.

Proof. Let B be any basis of the operator group T over $Z(T)$. Then it is a consequence of Theorems 1.2 and 6.1 that B is a basis of the operator group L over F . It is a consequence of Theorem 6.1 and of (4.5) that there exists an element q in F so that the n elements $q^{g'}$ for g in \mathbf{G} form a basis of F over $Z(T)$; and these elements form by Theorem 6.1 a basis of L over T .

Suppose now that the above conditions are satisfied. Then there exist in B two different elements b_1 and b_2 and there exists in \mathbf{G} an element $v \neq 1$ so that $q^{1-v'}$ is not in $Z(T)$. That this is possible is clear since $\sum_g q^{g'} = q \sum_g q^{g'-1}$ is an element $\neq 0$ in $Z(T)$ and q is not in $Z(T)$. The elements $1, v$ do not exhaust \mathbf{G} . Put $w = qb_1 + q^{v'}b_2$ and denote by W the set of all the elements: $t + zw$ for t in T and z in $Z(T)$. If g is an element in \mathbf{G} which is different from 1 , then $w^g = q^{g'}b_1 + q^{v'g'}b_2 \neq w$, since $q \neq q^{g'}$, and since b_1, b_2 are independent over F . Hence $T < W \leq L$.

It is obvious that W contains the sums of any two of its elements and that $Z(T) \leq Z(W)$. Suppose now that $t + zw = r \neq 0$ and that f be an element in F so that fr is in W . Then there exists an element s in T and an element h in $Z(T)$ so that

$$s + hqb_1 + hq^v b_2 = fr = ft + fzb_1 + fzq^v b_2$$

or

$$(\text{---}) \quad ft - s = (h - fz)qb_1 + (h - fz)q^v b_2.$$

Suppose now that f is not in $Z(T)$.

Case 1. $ft - s = 0$. Then it follows from the properties of T and from the fact that both s and t are in T that $t = 0$. Hence $r = zw$ and $r \neq 0$ implies $z \neq 0$. Consequently $h - fz \neq 0$ and therefore $qb_1 + q^v b_2 = 0$. But this is impossible.

Case 2. $ft - s \neq 0$. Then $h - fz \neq 0$. Suppose first that $t = 0$. Then we find for any element, g in G :

$$(h - f^g z)(q^g b_1 + q^v g^v b_2) = -s = (h - fz)(qb_1 + q^v b_2),$$

since s is in T and h, z are in $Z(T)$. Then

$$(h - f^g z)q^g = (h - fz)q \quad \text{and} \quad (h - f^g z)q^v g^v = (h - fz)q^v$$

and from $h - fz \neq 0$, it follows that $h - f^g z = (h - fz)g^v \neq 0$ so that $q^g q^v = q^v g^v q$ or $q^{1-v} = (q^{1-v})g^v$ for every g^v in G^v . Hence q^{1-v} is an element in $Z(T)$ and this is impossible by our choice of v , so that $t \neq 0$.

Every element x in T has the form: $x = \sum_{b \text{ in } B} z(x, b)b$ for $z(x, b)$ in $Z(T)$.

Then it follows from (---) that

$$\begin{aligned} fz(t, b) - z(s, b) &= 0 \quad \text{for } b \neq b_i, \\ fz(t, b_1) - z(s, b_1) &= (h - fz)q, \\ fz(t, b_2) - z(s, b_2) &= (h - fz)q^v, \end{aligned}$$

since the elements b in B form a basis of the operator group L over F . Since f is not in $Z(T)$, we find $z(t, b) = z(s, b) = 0$ for $b \neq b_i$. Since $t \neq 0$, this implies that not both $z(t, b_i)$ are 0. Eliminating f from the remaining two equations we find that

$$\begin{aligned} f[z(t, b_1) + zq] &= z(s, b_1) + hq, \\ f[z(t, b_2) + zq^v] &= z(s, b_2) + hq^v \quad \text{and therefore} \\ z[z z(s, b_2) - h z(t, b_2)] + q^v[h z(t, b_1) - z z(s, b_1)] \\ &= z(t, b_1)z(s, b_2) + z(t, b_2)z(s, b_1). \end{aligned}$$

Since the right side of this last equation is invariant under all the g^v in G^v , and since there are g^v in G^v which are different from both 1 and v^v , it follows now that

$$z z(s, b_2) = h z(t, b_2) \quad \text{and} \quad z z(s, b_1) = h z(t, b_1);$$

and consequently $z \neq 0$, since not both $z(t, b_i)$ are 0. Since all the $z(t, b)$ and $z(s, b)$ for $b \neq b_i$ are 0, it follows that $zs = ht$ and therefore we find from (—) that

$$(h - fz)[qb_1 + q^{v'}b_2] = ft - s = z^{-1}(fz - h)t$$

or that $-z^{-1}t = qb_1 + q^{v'}b_2$, since $h - fz \neq 0$. But this is impossible, since $z^{-1}t$ belongs to T and $w = qb_1 + q^{v'}b_2$ does not. Hence it is impossible that f is not in $Z(T)$; and this shows that W satisfies (a) to (c).

Assume now conversely that W is a domain between T and L which satisfies (b) and (c). To prove the necessity of our conditions we have to discuss two cases:

Case I. The order of \mathbf{G} is ≤ 2 . Since there is nothing to prove, if $\mathbf{G} = 1$, we suppose that \mathbf{G} consists of two different elements 1 and g . If w is any element in W , then $w = t_1q + t_2q^{g'}$ with t_i in T . Since T contains $t = t_1(q + q^{g'})$, W contains $w - t = (t_2 - t_1)q^{g'}$. Since $t_2 - t_1$ is in T , and since $q^{g'}$ does not belong to $Z(T)$, it follows from (b), (c) that $t_1 = t_2$ so that $w = t$ is an element in T , i. e. $W = T$.

Case II. The rank of the operator group L over F is 1. Then let t be any element $\neq 0$ in T . If w is an element in W , then $w = ft$ for some f in F and it follows from (c), (b) that f is in $Z(T)$ so that w is in T , i. e. $W = T$.

8. There exists a comparatively complete extension of the Galois Theory of finite groups of automorphisms (of commutative fields) to groups of automorphisms \mathbf{G} which satisfy the condition:

(F) The set of elements f^g for g in \mathbf{G} is finite for every f . The theory of these groups may be described as follows.⁷

(8.1) Let K be a subfield of the commutative field F . Then there exists a group \mathbf{G} of automorphisms of F so that

(1) the set $f^{\mathbf{G}}$ (of all the elements f^g for g in \mathbf{G}) is finite for every element f in F ,

(2)
$$K = (F, \mathbf{G})$$

if, and only if, F is algebraic, normal and separable over K .

(8.2) If F is algebraic, normal and separable over its subfield K , then

(a) conditions (1), (2) of (8.1) are satisfied by $(K < F)$;

(b) F is algebraic, normal and separable over every field between K and F ;

⁷ Cp. footnote ³.

(c) every K -automorphism of a field between K and F is induced by an automorphism of F ;

(d) every finite set of elements in F is contained in a field between K and F that is finite, normal and separable over K .

The groups G which satisfy condition (1) of (8.1) and $G = ((F, G) < F)$ may be characterized by a certain closure property which we need not state, as we are not going to make any use of it. The following result is however of some importance for us.

(8.3) Every subgroup S of the group G of automorphisms of F which has the property (1) of (8.1) satisfies $S = ((F, S) < F)$ if, and only if G is finite.

9. We shall now develop the Galois Theory of groups of automorphisms of linear systems which are subject to the above-mentioned condition (F) in analogy to the theories of sections 6 and 8.

THEOREM 9.1. *The subset T of the linear system L over the commutative field F satisfies:*

(a) *the identity is the only T -automorphism of L which induces the identity in F ,*

(b) *$x^{(T < L)}$ is a finite set for every element x in L ,*

(c)
$$T = (L, (T < L))$$
if, and only if,

(i) *T is complete in L ,*

(ii) *F is algebraic, normal and separable over $Z(T)$,*

(iii) *L is the direct product of F and T .*

Proof. Suppose first that the conditions (a) to (c) are satisfied by T . Then it is a consequence of Theorem 3.1 that T is complete in L , that $Z(T) = (F, (Z(T) < F))$ and that L is the direct product of F and T . There exists therefore an element $t \neq 0$ in T . If f is any element in F , then $(ft)^{(T < L)} = f^{(T < L)}t$ is a finite set so that $f^{(T < L)}$ is a finite set. Since L is the direct product of F and T , it follows from Theorem 2.3 that $(T < L)' = (Z(T) < F)$ so that finally every set $f^{(Z(T) < F)}$ for f in F is finite. Now it follows from (8.1) that F is algebraic, normal and separable over $Z(T)$ so that (i) to (iii) are consequences of (a) to (c).

If conversely the conditions (i) to (iii) are satisfied, then it follows from (8.1) that $Z(T) = (F, (Z(T) < F))$ and it follows therefore from Theorem

3.1 that $T = (L, (T < L))$ and that the identity is the only T -automorphism of L which induces the identity in F . Furthermore it follows from (8.1) that every set $f^{(T < L)'}$ for f in F is finite, since by condition (iii) and Theorem 2.3 we have $(T < L)' = (Z(T) < F)$. By (iii) there exist to every element x in L elements t_i in T , f_i in F so that $x = \sum_{i=1}^n t_i f_i$. Consequently $x^{(T < L)}$ is a subset of the set $\sum_{i=1}^n t_i f_i^{(T < L)'}$ which is finite so that (a) to (c) are consequences of (i) to (iii).

THEOREM 9.2. *The group \mathbf{G} of automorphisms of the linear system L over the commutative field F has the properties:*

(a) *the identity is the only (L, \mathbf{G}) -automorphism of L that induces the identity in F , and*

(b) *$x^{((L, \mathbf{G}) < L)}$ is a finite set for every element x in L*

if, and only if, the following conditions are satisfied by \mathbf{G} :

(i) *if \mathbf{S} is a normal subgroup of finite index in \mathbf{G} , and if \mathbf{S} is the cross cut of \mathbf{G} and $((L, \mathbf{S}) < L)$, then every automorphism in \mathbf{G} which induces the identity in $Z((L, \mathbf{S}))$ belongs to \mathbf{S} ;*

(ii) *$x^{\mathbf{G}}$ is a finite set for every element x in L .*

Proof. It is clear that (ii) is a consequence of (b), since $\mathbf{G} \leq ((L, \mathbf{G}) < L)$. If $T = (L, \mathbf{G})$, then it is a consequence of (a), (b) and of $(L, \mathbf{G}) = (L, ((L, \mathbf{G}) < L))$ that conditions (a) to (c) of Theorem 9.1 are satisfied so that T is complete in L , F is algebraic, normal and separable over $Z(T)$, and L is the direct product of F and T . If \mathbf{S} is any subgroup of \mathbf{G} and $B = (L, \mathbf{S})$, then $B = (L, (B < L))$ and it follows from Theorem 3.4 that B is complete in L and that $B = Z(B)T$. If \mathbf{g} is an automorphism in \mathbf{G} so that \mathbf{g}' is a $Z(B)$ -automorphism, then \mathbf{g} is a T -automorphism and therefore a B -automorphism of L so that \mathbf{g} belongs to the cross-cut of \mathbf{G} and $((L, \mathbf{S}) < L)$; and this contains (i) as a special case.

Suppose now that conditions (i) and (ii) are satisfied by the group \mathbf{G} . If u is any element $\neq 0$ in L , then denote by \mathbf{S} the set of all those automorphisms in \mathbf{G} which leave every element in the finite set $u^{\mathbf{G}}$ invariant. Then \mathbf{S} is a normal subgroup of finite index in \mathbf{G} ; and \mathbf{G}/\mathbf{S} is essentially the finite (transitive) group of permutations which \mathbf{G} induces in $u^{\mathbf{G}}$. Since $U = (L, \mathbf{S})$ contains $u^{\mathbf{G}}$, it follows that \mathbf{S} is the cross-cut of \mathbf{G} and $(U < L)$ —an automorphism, leaving every element in U invariant, has in particular every element in $u^{\mathbf{G}}$ as a fixed element—so that the conditions of (i) are satisfied

by S . Hence S contains every automorphism in G which induces the identity in $Z((L, S)) = Z(U)$.—Since $U \neq 0$ —for U contains $u \neq 0$ —it follows from (2.2) that U is complete in L and that $Z(U) = (F, S')$. G induces in the linear system U over $Z(U)$ a group G^* of automorphisms, since S is a normal subgroup of G , since $(L, S)^g = (L, g^{-1}Sg)$, and since therefore every automorphism g in G induces an automorphism g^* in U . If g is in G , and if $g^* = 1$, then g is in the cross-cut of G and $(U < L)$ and therefore in S . The group G^* of all the automorphisms g^* for g in G is therefore essentially the same as G/S so that G^* is in particular a finite group of automorphisms of the linear system U over $Z(U)$. If $g^{**} = 1$, then g' leaves every element in $Z(U)$ invariant so that—as has been remarked before— g belongs to S , i. e. $g^* = 1$. Finally $T = (L, G) \leq U = (L, S)$ and therefore $(U, G^*) = (L, G) = T$. Hence it follows from Theorem 6.1 that T is complete in U , U is the direct product of $Z(U)$ and T and $Z(U)$ is finite, normal and separable over $Z(T)$.

Special consequences of this last result—as applied to every u in L —are that T is complete in L and that $L = FT$; and this implies that (a) holds true.

If $t \neq 0$ is an element in T , f any element in F , then $(ft)^G = f^{G'}t$ is a finite set of elements in L ; and consequently $f^{G'}$ is a finite set of elements in F . Since $Z(T) = (F, G')$ by (2.2), it follows from (8.1) that F is algebraic, normal and separable over $Z(T)$ so that $Z(T) = (F, (Z(T) < F))$.

Suppose now that the elements b_1, \dots, b_k in F are independent over $Z(T)$. Denote by S the set of all the automorphisms g in G so that g' leaves every element in every set $b_i^{G'}$ invariant. Since all the sets $f^{G'}$ for f in F are finite, it follows that S satisfies all the conditions of (i). Hence it follows from what has been proved in the second paragraph of the proof that (L, S) is the direct product of $Z((L, S))$ and T , and $Z((L, S)) = (F, S')$ contains all the elements b_i . Since $T = (L, G) = ((L, S), G^*)$ —in the notation of the second paragraph of the proof—this implies that the b_i are independent over T too. Hence L is the direct product of F and T and now it follows from Theorem 9.1 that the condition (b) of our theorem is satisfied by G .⁸

Combining the results of Theorems 9.1 and 9.2 we find the following corollary which takes the place of Theorem 6.1 in this section.

⁸It might be worth noting that the conditions (i) to (iii) of Theorem 9.1 have been derived here from the conditions (i) and (ii) of this Theorem 9.2 without any recurrence to Theorem 3.1.

COROLLARY 9.3. *Let T be a subset of the linear system L over the field F . Then there exists a group \mathbf{G} of automorphisms of L which satisfies conditions*

(i) and (ii) of Theorem 9.2 and which satisfies:

$$T = (L, \mathbf{G})$$

if, and only if, T is complete in L , F is algebraic, normal and separable over $Z(T)$, and L is the direct product of F and T .

COROLLARY 9.4. *Suppose that the subset T of the linear system L over the commutative field F is complete in L , and that F is algebraic, normal and separable over $Z(T)$. Then L is the direct product of F and T if, and only if, every $Z(T)$ -automorphism of F is induced by one and only one T -automorphism of L .*

Proof. It is a consequence of Theorem 2.3, (a) that every $Z(T)$ -automorphism of F is induced by one and only one T -automorphism of L , if only L is the direct product of F and T . Assume conversely that every $Z(T)$ -automorphism of F is induced by one and only one T -automorphism of L . Then it is a consequence of Theorem 2.3, (b) that $L = FT$. Since $(T < L)' = (Z(T) < F)$, and since F is algebraic, normal and separable over $Z(T)$, it follows from (8.1) that $Z(T) = (F, (Z(T) < F)) = (F, (T < L)')$ and consequently it follows now from Theorem 2.3, (c) that every basis of the operator group T over $Z(T)$ is a basis of the operator group L over F ; and hence it follows from Theorem 1.2 that L is the direct product of F and T .

THEOREM 9.5. *A group \mathbf{G} of automorphisms of the linear system L over the commutative field F such that $x^{\mathbf{G}}$ is a finite set for every element x in L , and such that condition (i) of Theorem 9.2 is fulfilled by \mathbf{G} , satisfies*

$$\mathbf{G} = ((L, \mathbf{G}) < L) \text{ if, and only if, } \mathbf{G}' = ((F, \mathbf{G}') < F).$$

Proof. It is a consequence of Theorem 9.2 and of the conditions imposed on \mathbf{G} , that the identity is the only (L, \mathbf{G}) -automorphism of L which induces the identity in F and that every set $x^{((L, \mathbf{G}) < L)}$ is finite for every x in L . Hence it follows from Theorem 9.1 that $(L, \mathbf{G}) = T$ is complete in L , that F is algebraic, normal and separable over $Z(T)$ and that L is the direct product of F and T . It is a consequence of (2.2) that $Z(T) = (F, \mathbf{G}')$, and it is a consequence of Theorem 2.3, (a) that every (F, \mathbf{G}') -automorphism of F is induced by one and only one (L, \mathbf{G}) -automorphism of L . Now our theorem is a consequence of Theorem 3.2.

It may finally be mentioned that Theorem 6.4 may be extended to our

case with hardly any change. Another immediate consequence of the theorems of this section and of (8.3) is the following statement:

Suppose that L is a linear system over the commutative field F , and that the group G of automorphisms of L satisfies:

- (a) x^G is a finite set for every x in L ;
- (b) if S is a normal subgroup of finite index in G , and if S contains all the (L, S) -automorphisms in G , then S contains every automorphism in G which induces a $Z((L, S))$ -automorphism in F .

Then every subgroup T of G satisfies $T = ((L, T) < L)$ if, and only if, G is finite.

10. Applications of the theory of linear systems to the theory of rings and non-commutative fields shall be given in this section. If R is a ring, if the commutative field F is part of the central of R , and if R and F have the same identity, then it is clearly possible to consider R as a linear system over F , since this only means restricting one's attention to the addition in R and to the multiplication of elements in R by elements in F .

LEMMA 10.1. *If the field F is contained in the central of the ring R , and if S is a subring of R which contains the unit-element of F and R , then it is necessary and sufficient for the completeness of S in the linear system R over F that the cross-cut of F and S be a field.—If furthermore the linear system R over F is the direct product of F and S (in the sense of section 1), then every S -automorphism of the linear system R over F is at the same time an automorphism of the ring R .*

Proof. The first statement of the lemma is clear. If the linear system R over the field F is the direct product of F and of its subring S , then let B be a basis of F over the cross-cut $Z(S)$ of S and F ($Z(S)$ is a subfield of F). There exist to every element x in R uniquely determined elements $s(x, b)$ in S —all of which with a finite number of exceptions are 0—so that $x = \sum_{b \text{ in } B} bs(x, b)$.

If g is an automorphism of the linear system R over F , then g applied on F alone is an automorphism of the field F . Suppose now that g is an S -automorphism of the linear system R over F . Then

$$(xy)^g = \left[\sum_{b, d \text{ in } B} bds(x, b)s(y, d) \right]^g = \sum_{b, d \text{ in } B} b^g d^g s(x, b)s(y, d) = x^g y^g$$

and this completes the proof.

This lemma shows in particular that the automorphisms, constructed in Theorem 2.3, (a), are ring-automorphisms in our case.

As we are giving preference to the subfield F of the central of the ring R , we consider as automorphisms of R only such ring-automorphisms of R which map F upon itself, a hypothesis that will be satisfied for all the ring-automorphisms of R , in case we assume F to be the central of R .

Consequently we use the following notation: If \mathbf{G} is a group of automorphisms of the ring R , then \mathbf{G}' is the group of automorphisms which the automorphisms in \mathbf{G} induce in F . If T is a subset of R , then $(T < R)$ consists of all those automorphisms of the ring R which map F upon itself and leave every element in T invariant.

Now it has to be remarked that it is impossible to make use of Theorem 2.3, (b), since it may very well happen that there are no T -automorphisms $\neq 1$ of the ring R which induce the identity in F whereas there may exist T -automorphisms $\neq 1$ of the linear system R over F which induce the identity in F . On the other hand it is obvious that Theorem 2.3, (c) may be used, since ring-automorphisms are at the same time automorphisms of the linear system.

Let now T be a subring of R which contains the unit-element of F . No other subsets will be considered. Then an element f in F satisfies $fT \leq T$ if, and only if, f is in T too; and for this reason we denote by $Z(T)$ the cross-cut of F and T . T is complete in R if, and only if, $Z(T)$ is a subfield of F .

Now it is easy to derive the following statements from Theorems 6.1 and 6.3.

THEOREM A. *Suppose that the cross-cut $Z(T)$ of the subring T of the ring R and of the subfield F of the central of the ring R is a subfield of F . Then there exists a finite group \mathbf{G} of automorphisms of the ring R —all of which map F upon itself—such that the identity is the only F -automorphism in \mathbf{G} , and such that $T = (R, \mathbf{G})$ if, and only if, F is finite, normal and separable over $Z(T)$ and the linear system R over F is the direct product of T and F .*

THEOREM B. *If \mathbf{G} is a finite group of automorphisms of the ring R all of which map the subfield F of the central of R upon itself, and if the identity is the only F -automorphism in \mathbf{G} , then*

$$\mathbf{G} = ((R, \mathbf{G}) < R).$$

The statements we are going to derive from Corollary 9.3 and Theorem 9.5 concern groups \mathbf{G} of automorphisms of the ring R with the following properties:

- (1) $F = F^{\mathbf{g}}$ for every \mathbf{g} in \mathbf{G} ;

- (2) x^G is a finite set of elements for every x in R ;
 (3) if S is a normal subgroup of finite index in G , and if S contains all those automorphisms in G which leave all the elements in (R, S) invariant, then every $Z((R, S))$ -automorphism in G belongs to S .

Note that a finite group G satisfies these conditions, if its automorphisms map F upon itself, and if the identity is the only F -automorphism in G .

THEOREM A'. Suppose that F is a subfield of the central of the ring R , and that T is a subring of R whose cross-cut with F is a subfield $Z(T)$ of F . Then there exists a group G of automorphisms of the ring R which satisfies the above conditions (1) to (3) so that

$$T = (R, G)$$

if, and only if, F is algebraic, normal and separable over $Z(T)$, and the linear system R over F is the direct product of F and T .

THEOREM B'. If the group G of automorphisms of the ring R satisfies conditions (1) to (3), then $G' = ((F, G') < F)$ is a necessary and sufficient condition for $G = ((R, G) < R)$.

The following important and obvious consequence of Theorem A' may be stated for future reference.

LEMMA 10.2. Suppose that the central of the ring R is a field F , and that the group G of automorphisms of the ring R satisfies conditions (2) and (3).

- (a) F is the centralizer of (R, G) in R .
 (b) $Z((R, G))$ is the central of (R, G) .

(b) is a consequence of (a); and (a) follows from the fact that by Theorem A' the linear system R over F is equal to $F(R, G)$, and that F is exactly the central of R .

Finally it may be noted that the following statement may be derived from Theorem 3.4.

THEOREM C. Suppose that F is a subfield of the central of the ring R ; and that the group G of automorphisms of the ring R satisfies conditions (1) to (3).

- (a) The set B between (R, G) and R satisfies $B = (R, (B < R))$ if, and only if, there exists a field S between $Z((R, G))$ and F so that $B = S(R, G)$.
 (b) If S is a field between $Z((R, G))$ and F , then $S = Z(S(R, G))$.

It is the main-objective of this section to show that in case of (non-

commutative) fields it is possible to prove an essentially stronger theorem than Theorem C.

LEMMA 10.3. *If the central F of the ring R is a field, if G is a finite group of automorphisms of R such that the identity is the only automorphism in G which leaves every element in F invariant and such that (R, G) is a subfield of R , if W is a ring between (R, G) and R whose cross-cut with F is the same as the cross-cut of (R, G) and F , then $W = (R, G)$.*

Note that every automorphism of R maps F upon itself, since F is the central of R , and that (R, G) though a field need not be a commutative field.

Proof. It is a consequence of Theorem A that F is finite, normal and separable over the cross-cut $Z(K)$ of $K = (R, G)$ and F , and that the linear system R over the commutative field F is the direct product of F and K . There exists therefore by (4.5) an element b in F so that the elements b^g for g in G form a basis of F over $Z(K)$, since the elements in G induce in F an isomorphic group of automorphisms, and since $(F, G) = Z(K)$. There exist furthermore to every element x in R uniquely determined elements $c(x, g)$ in K so that $x = \sum_{g \text{ in } G} c(x, g)b^g$ and x belongs to K if, and only if, all the elements $c(x, g)$ are equal.

Assume now that $K < W$. Then there exist in W elements which are not contained in K ; and amongst these there is one, w , so that the number of coefficients $c(x, g) \neq 0$ is as small as possible. Since w is not in K , $w \neq 0$ and there exists an automorphism v in G so that $c(w, v) \neq 0$.

Let now t be any element in K . Then

$$twc(w, v)^{-1} - wc(w, v)^{-1}t = \sum_{g \text{ in } G} [tc(w, g)c(w, v)^{-1} - c(w, g)c(w, v)^{-1}t]b^g$$

would be in W and the number of its coefficients $\neq 0$ would be smaller than for w . Hence this element is in K so that all its coefficients are equal. Since at least one of these coefficients is 0, all the coefficients are 0 so that

$$tc(w, g)c(w, v)^{-1} = c(w, g)c(w, v^{-1})t \text{ for every } t \text{ in } K, g \text{ in } G.$$

Now it follows from Lemma 10.2 that $z(w, g) = c(w, g)c(w, v)^{-1}$ is an element in F , and since $z(w, g)$ is an element in K , it is in the cross-cut $Z(K)$ of K and F . Hence $w = \sum_{g \text{ in } G} z(w, g)b^g c(w, v) = f c(w, v)$. Since $w \neq 0$, f is an element in the cross-cut of W and F ; and it follows from our hypothesis that f is in $Z(K)$. Thus w would be in K and this is a contradiction so that finally $W = K$.

THEOREM 10.4. *If G is a finite group of automorphisms of the (non-commutative) field Q such that the identity is the only automorphism in G*

which leaves every element in the central F of Q invariant, then every ring R between (Q, \mathbf{G}) and Q whose cross-cut with F is a field satisfies:

$$R = (Q, (R < Q)).$$

Note that every (Q, \mathbf{H}) is a subfield of Q so that the condition imposed on R is necessary and sufficient and implies that R is a field.

Proof. Denote by $Z(R)$ the cross-cut of R and F . $Z(R)$ is a subfield of F which contains the cross-cut $Z(K)$ of $K = (Q, \mathbf{G})$ and F . Put $S = Z(R)K$. Then $K \leq S \leq R$; and it follows from Theorem A and Theorem C that $Z(S) = Z(R)$ and $S = (Q, (S < Q))$, since F is finite, normal and separable over $Z(K)$ and since therefore the field $Z(R)$ between $Z(K)$ and F satisfies: $Z(R) = (F, (Z(R) < F))$. This implies in particular that S is a field. Since $(S < Q) \leq \mathbf{G}$, it follows now from Lemma 10.3 that $S = R$ and this completes the proof.

THEOREM 10.5. *If K is a subfield of the field Q such that the identity is the only K -automorphism of Q which leaves every element in the central F of Q invariant, such that the sets $x^{(K < Q)}$ are finite for every element x in Q , and $K = (Q, (K < Q))$, then every ring R between K and Q whose cross-cut with F is a field satisfies:*

$$R = (Q, (R < Q)).$$

Proof. It is a consequence of Theorem 9.1 that F is algebraic, normal and separable over the cross-cut $Z(K)$ of K and F , and that the linear system Q over F is the direct product of F and K i.e. the group $(K < Q)$ satisfies the conditions (1) to (3). If the cross-cut $Z(R)$ of the ring R between K and Q is a field, then $Z(R)$ is a field between $Z(K)$ and F ; and it follows from (8.1) and (8.2) that $(F, (Z(R) < F)) = Z(R)$. It is then a consequence of Theorem C that the domain $S = Z(R)K$ satisfies: $S = (Q, (S < Q))$ and $Z(S) = Z(R)$. Since $(S < Q) \leq (K < Q)$ it follows that the identity is the only S -automorphism of Q which induces the identity in F and that every set $x^{(S < Q)}$ is finite. Finally it is clear that S is a field which is contained in R .

Suppose now that u is any element in R . Denote by \mathbf{U} the set of all the S -automorphisms of Q which leave all the elements in $u^{(S < Q)}$ invariant and put $V = (Q, \mathbf{U})$. It is clear that \mathbf{U} is a normal subgroup of finite index in $(S < Q)$, that $u^{(S < Q)} \leq V$ and that therefore $\mathbf{U} = ((Q, \mathbf{U}) < Q)$. It is a consequence of Theorem 9.2 that an S -automorphism of Q which induces the identity in $R(V)$ leaves every element in V invariant. Thus $(S < Q)$ induces in the field V with central $R(V)$ a finite group \mathbf{G} of automorphisms so that the identity is the only automorphism in \mathbf{G} which induces the identity in

$Z(V)$ and so that $(V, \mathbf{G}) = S$. Denote now by D the cross-cut of V and R . D is a ring between S and V which contains u and whose cross-cut with $Z(V)$ is just $Z(R) = Z(S)$. Hence it follows from Lemma 10.3 that $D = S$ so that in particular u is an element in S . Hence $S = R$ and this completes the proof.

CHAPTER III. Crossed Products.⁹

11. The extension of the concept of crossed product we are going to give here concerns itself with a (not necessarily commutative) field Q whose central may be denoted by F and a group \mathbf{G} of automorphisms of the field Q which is subject to the following conditions:

(I) Q is the direct product of F and the subfield $K = (Q, \mathbf{G})$ of Q ;

(II) $K = (Q, (K < Q))$.

Two important inferences of (I) may be stated at once.

(I') The identity is the only F -automorphism in \mathbf{G} .

(I'') F is the centralizer of K in Q ; and the central of K is its cross-cut Z with F .

Given condition (I), one verifies that (II) is equivalent to the following condition:

(II') $Z = (F, (Z < F))$.

It may be noted furthermore that a consequence of (I) is

(I*) Every Z -automorphism of F is induced by one and only one K -automorphism of Q .

Upon occasion we shall have to use the further restriction:

(III) $\mathbf{G} = (K < Q)$.

In denoting by g' the automorphism of F which is induced by g in Q , it is a consequence of (I*) that (III) is equivalent to the following assumption.

(III') $\mathbf{G}' = (Z < F)$.

Conditions (I) to (III) are satisfied by all those finite groups \mathbf{G} of automorphisms of Q whose only F -automorphism is the identity.—Conditions (I) and (II) are satisfied by the more general class of groups which satisfy the conditions (2), (3) stated in section 10.

Now we connect with every element g in \mathbf{G} an indeterminate $u(g)$ and consider the system $Q\mathbf{G}$ of all the linear forms:

⁹ For a presentation of the classical theory of crossed products cp. e. g. H. Hasse, *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 171-214.

$$\sum_{g \in G} u(g)q(g)$$

where the $q(g)$ are elements in Q all but a finite number of which are 0. It is clear how to add two such forms and how to multiply them by elements in Q [from the right].

In this linear system QG over Q a multiplication shall be defined which is subject to the following rules:

- (A) $qu(g) = u(g)q^*$ for q in Q and g in G ;
 (F) if g and h are two elements in G , then there exists an element (g, h) in Q so that $u(g)u(h) = u(gh)(g, h)$.

The elements (g, h) are called a *factor-set* and the linear system QG enriched by this multiplication is termed the *crossed-product*

$$(Q, G, (g, h)).$$

(11.1) The multiplication in $(Q, G, (g, h))$ is associative if, and only if,

- (i) every (g, h) is in F ;
 (ii) $(r, st)(s, t) = (rs, t)(r, s)^t$ for r, s, t in G .

Proof. Suppose first that the multiplication is associative. If q is any element in Q and g, h are elements in G , then

$$\begin{aligned} u(gh)(g, h)q &= u(g)u(h)q = q^{(gh)^{-1}}u(g)u(h) = q^{(gh)^{-1}}u(gh)(g, h) \\ &= u(gh)q(g, h) \text{ or} \end{aligned}$$

$(g, h)q = q(g, h)$ for every q in Q so that (i) holds true.

If furthermore r, s, t are three elements in G , then

$$\begin{aligned} u(rst)(r, st)(s, t) &= u(r)u(st)(s, t) = u(r)u(s)u(t) \\ &= u(rs)(r, s)u(t) = u(rs)u(t)(r, s)^t \\ &= u(rst)(rs, t)(r, s)^t \end{aligned}$$

and this proves the necessity of (ii).

If conversely (i) and (ii) are satisfied, then

$$\begin{aligned} \sum_r u(r)a(r) \left[\sum_s u(s)b(s) \sum_t u(t)c(t) \right] \\ &= \sum_{r, s, t} u(r)a(r)[u(s)b(s)u(t)c(t)] \\ &= \sum_{r, s, t} u(r)a(r)[u(s)u(t)b(s)^t c(t)] \\ &= \sum_{r, s, t} u(r)a(r)u(st)(s, t)b(s)^t c(t) \\ &= \sum_{r, s, t} u(r)u(st)a(r)^{st}(s, t)b(s)^t c(t) \\ &= \sum_{r, s, t} u(rst)(r, st)(s, t)a(r)^{st}b(s)^t c(t) \\ &= \sum_{r, s, t} u(rst)(rs, t)(r, s)^t a(r)^{st}b(s)^t c(t) \end{aligned}$$

$$\begin{aligned}
&= \sum_{r,s,t} u(rs)u(t)(r,s)^t a(r)^{st} b(s)^t c(t) \\
&= \sum_{r,s,t} u(rs)(r,s)[u(t)a(r)^{st} b(s)^t c(t)] \\
&= \sum_{r,s,t} [u(r)u(s)][u(t)a(r)^{st} b(s)^t c(t)] \\
&= \sum_{r,s,t} [u(r)u(s)a(r)^{st} b(s)][u(t)c(t)] \\
&= \sum_{r,s,t} [u(r)a(r)u(s)b(s)][u(t)c(t)] \\
&= \left[\sum_r u(r)a(r) \sum_s u(s)b(s) \right] \sum_t u(t)c(t)
\end{aligned}$$

and this completes the proof.

This statement explains why we have to and are going to restrict ourselves to the consideration of factor-sets which satisfy the above conditions (i) and (ii).

As one verifies easily that the element $u(1)(1,1)^{-1}$ is the unit element in $(Q, G, (g, h))$, we may assume without loss of generality that

$$(iii) \quad (g, 1) = (1, h) = 1 \text{ or } u(1) = 1.$$

Finally one verifies that $u(g^{-1})(g, g^{-1})^{-1}$ is the inverse to $u(g)$.

12. In this section we discuss the general structural properties of crossed products

$$P = (Q, G, (g, h))$$

where Q is a field, G a group of automorphisms of Q which satisfies (I) and (II), and where (g, h) is a factor-set of G in Q which satisfies (i) to (iii). (12.1) P is simple.

Proof. Suppose that W is a two sided ideal $\neq 0$ in P . Then there exists among the elements $w = \sum_{g \in G} u(g)q(w, g) \neq 0$ in W at least one such that the number of g with $q(w, g) \neq 0$ is as small as possible. Let v be such an element, and suppose that u is an automorphism with $q(v, u) \neq 0$. If g is another automorphism in G , and if $u \neq g$, then it is a consequence of (I*) (in section 11) that $u' \neq g'$ and there exists therefore an element f in F so that $f^u \neq f^g$. Clearly $fv - vf^u = \sum_{h \in G} u(h)[(f^h - f^u)q(v, h)]$ is an element in W ; and it is 0, since the number of its coefficients $\neq 0$ is smaller than for v . Hence in particular $(f^g - f^u)q(v, g) = 0$ and this implies $q(v, g) = 0$ for every $g \neq u$. This implies that $u(u)$ itself is an element in W ; and hence all the $u(g)$ are in W , i. e. $P = W$.

(12.2) $(Q, G, (g, h))$ is the direct product of $(F, G', (g, h))$ and K .

This is an obvious consequence of condition (I) in section 11.

An interesting consequence of this statement and of a well-known property

of crossed-products of commutative fields by finite groups of automorphisms may be stated separately.

(12.2*) If \mathbf{G} is a finite group of automorphisms of the field Q so that the identity is the only central-automorphism in \mathbf{G} , then $(Q, \mathbf{G}, (\mathbf{g}, \mathbf{h}) = 1)$ is a full matrix-algebra over the field (Q, \mathbf{G}) .

(12.3) An element w in $P = (Q, \mathbf{G}, (\mathbf{g}, \mathbf{h}))$ satisfies $wF = Fw$ if, and only if, it has the form $u(\mathbf{g})q$ for some \mathbf{g} in \mathbf{G} and q in Q .

Proof. That elements of the form $w = u(\mathbf{g})q$ satisfy $wF = Fw$, is obvious. If on the other hand $w \neq 0$ satisfies $wF = Fw$, then there exists to every element f in F an element f^* in F so that $fw = wf^*$. If $w = \sum_{\mathbf{g} \in \mathbf{G}} u(\mathbf{g})q(\mathbf{w}, \mathbf{g})$, then this implies that $f^*q(\mathbf{w}, \mathbf{g}) = f^*q(\mathbf{w}, \mathbf{g})$ for every \mathbf{g} in \mathbf{G} . If \mathbf{u} and \mathbf{v} are two different automorphisms in \mathbf{G} so that both $q(\mathbf{w}, \mathbf{u})$ and $q(\mathbf{w}, \mathbf{v})$ are different from 0, then this would imply that $f^{\mathbf{u}} = f^{\mathbf{v}}$ for every f in F ; and this is impossible by (I*) of section 11. Hence $w = u(\mathbf{g})q$.

(12.4) Q is the centralizer of F in P .

If w is an element, satisfying $wf = fw$ for every f in F , then it follows from (12.3) that $w = u(\mathbf{g})q$ and it follows from (I*) that $\mathbf{g} = 1$.

(12.5) Q is uniquely determined as the greatest subfield of P which is contained in the normalizer of F in P .

Proof. Suppose that U is some subfield of the normalizer of F in P . Then it follows from (12.3) that every element in U has the form $u(\mathbf{g})q$. If $u(\mathbf{g})q$ and $u(\mathbf{h})p$ are two elements in U which are both different from 0, then their sum is in U and therefore of the one-term-form, i. e. $\mathbf{g} = \mathbf{h}$. For the same reason $\mathbf{g} = \mathbf{g}^2$, i. e. $\mathbf{g} = 1$ so that $U \leq Q$.

(12.6) Z is the central of P .

For elements of the central belong to Q by (12.4), hence to F . They belong to K and therefore to Z , since they permute with the elements $u(\mathbf{g})$.

(12.7) K is the centralizer of $(F, \mathbf{G}', (\mathbf{g}, \mathbf{h}))$.

This follows from (12.4), since the $u(\mathbf{g})$ are in $(F, \mathbf{G}', (\mathbf{g}, \mathbf{h}))$.

13. It is a consequence of (12.3) that the normalizer of both the fields F and Q in $P = (Q, \mathbf{G}, (\mathbf{g}, \mathbf{h}))$ is—apart from 0—the group, generated in adjoining the elements $u(\mathbf{g})$ to Q . Denote the set of all the elements w in P which satisfy: $wF = Fw$ —or $wQ = Qw$ —by N ; so that the elements $\neq 0$ in N form the group, we have described just now.

Every automorphism of P maps the central $Z = (F, \mathbf{G}')$ upon itself. But there may exist automorphisms of P which do not map F upon itself.

If however an automorphism of P maps F upon itself, then it maps Q and N upon themselves. If conversely an automorphism of P maps N upon itself, then it follows from (12.5) that this automorphism maps Q upon itself; and if an automorphism maps Q upon itself, then F is mapped upon itself too, since F is the central of Q . In this section we shall investigate *those automorphisms of P which map F , Q and N upon themselves.*

If the automorphism \mathbf{r} of P maps F , Q and N upon themselves, then \mathbf{r} induces an automorphism \mathbf{r}^* in Q ; and \mathbf{r}^* —in the usual notation—is the automorphism which \mathbf{r} and \mathbf{r}^* induce in F . Since \mathbf{r} induces an automorphism in the group N^* which maps the cross-cut Q^* of Q and N^* upon itself, it follows that \mathbf{r} induces an automorphism in the quotient group N^*/Q^* , and since \mathbf{G} and N^*/Q^* are essentially the same, it follows that \mathbf{r} induces an automorphism \mathbf{r}'' in \mathbf{G} , and that

$$(a) \quad u(\mathbf{g})^{\mathbf{r}} = u(\mathbf{g}^{\mathbf{r}''})\mathbf{r}(\mathbf{g}) \text{ with } \mathbf{r}(\mathbf{g}) \text{ in } Q^*.$$

Applying now \mathbf{r} upon condition (A) of section 11, we find that

$$\begin{aligned} u(\mathbf{g}^{\mathbf{r}''})q^{\mathbf{r}^*\mathbf{g}^{\mathbf{r}''}}\mathbf{r}(\mathbf{g}) &= q^{\mathbf{r}^*}u(\mathbf{g}^{\mathbf{r}''})\mathbf{r}(\mathbf{g}) = q^{\mathbf{r}^*}u(\mathbf{g})^{\mathbf{r}} = (qu(\mathbf{g}))^{\mathbf{r}} \\ &= (u(\mathbf{g})q^{\mathbf{g}})^{\mathbf{r}} = u(\mathbf{g}^{\mathbf{r}''})\mathbf{r}(\mathbf{g})q^{\mathbf{g}^{\mathbf{r}^*}} \end{aligned}$$

or

$$(b) \quad q^{\mathbf{r}^*\mathbf{g}^{\mathbf{r}''}}\mathbf{r}(\mathbf{g}) = \mathbf{r}(\mathbf{g})q^{\mathbf{g}^{\mathbf{r}^*}} \text{ for } q \text{ in } Q \text{ and } \mathbf{g} \text{ in } \mathbf{G}$$

or what amounts to the same

$$(b') \quad q = \mathbf{r}(\mathbf{g})^{-1}q^{\mathbf{r}^*\mathbf{g}^{\mathbf{r}''}}\mathbf{r}(\mathbf{g}).$$

Applying the automorphism \mathbf{r} upon condition (F) of section 11, and in using conditions (i), (ii) of (11.1), it follows that

$$\begin{aligned} (\mathbf{g}, \mathbf{h})^{\mathbf{r}^*} &= [u(\mathbf{gh})^{-1}u(\mathbf{g})u(\mathbf{h})]^{\mathbf{r}} = \mathbf{r}(\mathbf{gh})^{-1}u(\mathbf{g}^{\mathbf{r}''}\mathbf{h}^{\mathbf{r}''})^{-1}u(\mathbf{g}^{\mathbf{r}''})\mathbf{r}(\mathbf{g})u(\mathbf{h}^{\mathbf{r}''})\mathbf{r}(\mathbf{h}) \\ &= \mathbf{r}(\mathbf{gh})^{-1}u(\mathbf{g}^{\mathbf{r}''}\mathbf{h}^{\mathbf{r}''})^{-1}u(\mathbf{g}^{\mathbf{r}''})u(\mathbf{h}^{\mathbf{r}''})\mathbf{r}(\mathbf{g})^{\mathbf{h}^{\mathbf{r}''}}\mathbf{r}(\mathbf{h}) \\ &= \mathbf{r}(\mathbf{gh})^{-1}(\mathbf{g}^{\mathbf{r}''}, \mathbf{h}^{\mathbf{r}''})\mathbf{r}(\mathbf{g})^{\mathbf{h}^{\mathbf{r}''}}\mathbf{r}(\mathbf{h}) \text{ or} \end{aligned}$$

$$(c) \quad (\mathbf{g}, \mathbf{h})^{\mathbf{r}^*}(\mathbf{g}^{\mathbf{r}''}, \mathbf{h}^{\mathbf{r}''})^{-1} = \mathbf{r}(\mathbf{gh})^{-1}\mathbf{r}(\mathbf{g})^{\mathbf{h}^{\mathbf{r}''}}\mathbf{r}(\mathbf{h}) \text{ for } \mathbf{g}, \mathbf{h} \text{ in } \mathbf{G}.$$

Thus we have seen that every automorphism \mathbf{r} of P which maps F , Q and N upon themselves induces an automorphism \mathbf{r}^* of Q , an automorphism \mathbf{r}'' of \mathbf{G} and—by (a)—a Q -valued function $\mathbf{r}(\mathbf{g})$ of the elements in \mathbf{G} ; and it is obvious that \mathbf{r} is uniquely determined by \mathbf{r}^* , \mathbf{r}'' and $\mathbf{r}(\mathbf{g})$.

THEOREM 13.1. *Suppose that \mathbf{r}^* is an automorphism of Q , \mathbf{r}'' an automorphism of \mathbf{G} , and that $\mathbf{r}(\mathbf{g})$ is a Q -valued function of the elements \mathbf{g} in \mathbf{G} . Then there exists an automorphism \mathbf{r} of P which induces \mathbf{r}^* in Q , \mathbf{r}'' in \mathbf{G} and satisfies (a) if, and only if, \mathbf{r}^* , \mathbf{r}'' and $\mathbf{r}(\mathbf{g})$ obey the rules (b), (c).*

Proof. The necessity of these conditions has already been verified.—

Thus assume that (b) and (c) are satisfied by r^* , r'' and $r(g)$. If x is any element in P , then there exist uniquely determined elements $q(x, g)$ —all of which with a finite number of exceptions are 0—so that

$$x = \sum_{g \in G} u(g) q(x, g).$$

A transformation r of P may be defined by

$$(d) \quad x^r = \sum_{g \in G} u(g^{r''}) r(g) q(x, g)^{r^*}$$

and this transformation satisfies clearly (a), induces r^* in Q , since $u(1) = 1$; and it will be clear that r induces r'' in G , as soon as we have proved that r is an automorphism of P . This transformation is a one-one-correspondence, mapping P upon the whole set P , since the equation $y^r = x$ possesses one and only one solution y in P , namely the element y with the coefficients $q(y, g) = r(g)^{-r^*-1} q(x, g^{r''}) r^{*-1}$. That r preserves addition is clear; that it preserves multiplication is verified as follows:

$$\begin{aligned} (xy)^r &= \left[\sum_{g,h} u(g) q(x, g) u(h) q(y, h) \right]^r = \left[\sum_{g,h} u(g) u(h) q(x, g)^h q(y, h) \right]^r \\ &= \left[\sum_{g,h} u(gh) (g, h) q(x, g)^h q(y, h) \right]^r \\ &= \sum_{g,h} u(g^{r''} h^{r''}) r(gh) (g, h)^{r^*} q(x, g)^{hr^*} q(y, h)^{r^*} \\ &= \sum_{g,h} u(g^{r''}) u(h^{r''}) (g^{r''}, h^{r''})^{-1} r(gh) (g, h)^{r^*} q(x, g)^{hr^*} q(y, h)^{r^*} \\ &= \sum_{g,h} u(g^{r''}) u(h^{r''}) r(g)^{h''} r(h) q(x, g)^{hr^*} q(y, h)^{r^*} \\ &= \sum_{g,h} u(g^{r''}) r(g) u(h^{r''}) r(h) q(x, g)^{hr^*} q(y, h)^{r^*} \\ &= \sum_{g,h} u(g^{r''}) r(g) u(h^{r''}) q(x, g)^{r^* h^{r''}} r(h) q(y, h)^{r^*} \\ &= \sum_{g,h} u(g^{r''}) r(g) q(x, g)^{r^*} u(h^{r''}) r(h) q(y, h)^{r^*} = x^r y^r \end{aligned}$$

and this completes the proof.

Restricting (b) to elements in F only we find

$$(b'') \quad r^{*'} g^{r''} = g' r^{*'}$$

and in applying (c) on $g = h = 1$ we find that

$$(c') \quad r(1) = 1$$

and in applying (c) on $h = g^{-1}$ we derive from (c') that

$$(c'') \quad (g, g^{-1}) (g^{r''}, g^{-r''})^{-1} = r(g)^{g^{r''}} r(g^{-1}).$$

14. In this section we add successively new hypotheses to those used in the preceding sections. To the hypothesis that r^* , r'' and $r(g)$ satisfy the conditions (b), (c) of section 13 we add first:

(1) $r^{*'}$ is an element in G' .

We note first that this assumption is certainly satisfied, whenever r^* is a Z -automorphism of Q and \mathbf{G} satisfies condition (III) of section 11.

From (1) it follows that there exists an automorphism w in \mathbf{G} so that $w' = r^{*'}.$ Then it follows from (b'') that

$$g'r'' = r^{*'}{}^{-1}g'r^{*'} = w'^{-1}g'w' = (w^{-1}gw)'$$

and this implies by (I*) of section 11 that

$$(1') \quad g'' = w^{-1}gw.$$

Let now be $s^* = r^*w^{-1}.$ Then it follows from (b) that

$$(1'') \quad \begin{aligned} q^{s^*gw}r(g) &= r(g)q^{gs^*w} \text{ or} \\ q^{s^*gr}(g)^{w^{-1}} &= r(g)^{w^{-1}}q^{gs^*}. \end{aligned}$$

Now we add another hypothesis.

(2) Every F -automorphism of Q is an inner automorphism of Q .

It is known that this hypothesis is a consequence of the finiteness of Q over F , and that this hypothesis is not always satisfied.

Since $s^{*'} = 1$, from (2) follows the existence of an element b in Q so that

$$q^{s^*} = b^{-1}qb \text{ for every } q \text{ in } Q.$$

Applying this on (1'') we find

$$q^g = [br(g)^{-w^{-1}b^{-g}}]^{-1}q^g[br(g)^{-w^{-1}b^{-g}}]$$

and it is a consequence of the fact that F is the central of Q that $br(g)^{-w^{-1}b^{-g}}$ is an element in F . Hence there exists an element $f(g)$ in F so that

$$(2') \quad r(g) = f(g)b^{(-g+1)w} = f(g)b^{-gw}bw$$

and it is a consequence of (c) that this F -valued function $f(g)$ satisfies

$$(2'') \quad (g, h)^w(w^{-1}gw, w^{-1}hw)^{-1} = f(gh)^{-1}f(g)^{w^{-1}hw}f(h).$$

If the F -valued function $f(g)$ satisfies condition (2''), then the identity-automorphism of Q together with the inner automorphism which is induced in \mathbf{G} by w together with this function $f(g)$ satisfy the conditions of the Theorem 13.1 so that they are induced by an automorphism of P which is a Q -automorphism of P and therefore a central-automorphism of P .

If we now add the final hypothesis that

(3) F -automorphisms of $(F, \mathbf{G}', (g, h))$ are inner automorphisms then it follows from the existence of an automorphism of P which leaves all the elements in Q invariant, induces in \mathbf{G} the inner automorphism effected by w , and induces $f(g)$ according to (a) of section 13, that there exists an element

in $(F, G', (g, h))$ which induces this automorphism. This element has by necessity the form $u(w)f$ for f in F , and now it follows that

$$\begin{aligned} u(w^{-1}gw)f(g) &= f^{-1}u(w)^{-1}u(g)u(w)f \\ &= f^{-1}u(w)^{-1}u(w^{-1})^{-1}u(w^{-1})u(g)u(w)f \\ &= f^{-1}(w^{-1}, w)^{-1}u(w^{-1}gw)(w^{-1}, gw)(g, w)f \\ &= u(w^{-1}gw)f^{1-w^{-1}gw}(w^{-1}, w)^{-w^{-1}gw}(w^{-1}, gw)(g, w) \end{aligned}$$

or

$$(3') \quad f(g) = f^{1-w^{-1}gw}(g/w) \text{ for } (g/w) = (w^{-1}, w)^{-w^{-1}gw}(w^{-1}, gw)(g, w)$$

and the function $r(g)$ has by (2') and (3') the form

$$r(g) = (b^w f)^{-w^{-1}gw+1}(g/w).$$

The most important special case of all these considerations may be stated separately.

If the field Q is finite over its central F , if G is a finite group of automorphisms of Q so that the identity is the only F -automorphism in G , if (g, h) is a factor-set of G in F which satisfies conditions (ii) and (iii) of section 11, if r^* is an (F, G') -automorphism of Q , r'' an automorphism of G and $r(g)$ a Q -valued function of the elements in G so that

$$\begin{aligned} q^{r^*gr''}r(g) &= r(g)q^{gr''} \text{ for } q \text{ in } Q, g \text{ in } G, \text{ and} \\ (g, h)^{r^*}(gr'', hr'')^{-1} &= r(gh)^{-1}r(g)^{hr''}r(h) \text{ for } g, h \text{ in } G, \end{aligned}$$

then there exists an element w in G and an element v in Q so that

$$\begin{aligned} w' &= r^{*'}, gr'' = w^{-1}gw \text{ and} \\ r(g) &= v^{-w^{-1}gw+1}(g/w) \end{aligned}$$

where

$$(g/w) = (w^{-1}, w)^{-w^{-1}gw}(w^{-1}, gw)(g, w).$$

If we choose in particular the factor-set $(g, h) = 1$, then we see:

If w is an automorphism in G and $r(g)$ a Q -valued function so that

$$r(gh) = r(g)^{w^{-1}hw}r(h),$$

then there exists an element v in Q so that

$$r(g) = v^{-w^{-1}gw+1}.$$

Finally it ought to be mentioned that the element v induces in Q the same automorphism as r^*w^{-1} .

THE NUMBER OF REPRESENTATIONS FUNCTION FOR BINARY QUADRATIC FORMS.*

By NEWMAN A. HALL.¹

The problem of finding the number of representations of an arbitrary integer by a given binary quadratic form has yet to be solved in complete generality. In the two centuries that have followed the first general investigation by J. L. Lagrange² of any part of the problem, the investigations have proceeded in two directions. A great number of specific forms have been considered individually for which more or less general solutions have been given. Again, certain general investigations have reduced the problem to more simple and direct questions. The early investigations of Dirichlet³ and more recently those of Pall⁴ are of this nature.

In the discussion to follow, we offer as a contribution to the general problem, the general explicit expressions for the number of representations function for all forms whose discriminant is such that there is a single class in each genus together with a specific example showing the numerical computation of the number of representations.

We are concerned with binary quadratic forms designated by $[a, b, c]$, of discriminant $-\Delta = b^2 - 4ac$, and shall examine the form of the number of representations function

$$N[m = ax^2 + bxy + cy^2]$$

this being the number of solutions in integers, x and y , of

$$m = ax^2 + bxy + cy^2.$$

As is customary only forms which are positive definite and whose coefficients have no common factors, i. e. are primitive, will be considered.

We shall base the investigation on the following theorem of Dirichlet:⁵

* Received September 20, 1939.

¹ Presented to the American Mathematical Society September 6, 1938, cf. *Bulletin of the American Mathematical Society*, vol. 44 (1938), p. 488.

² J. L. Lagrange, "Recherches d'Arithmetique," *Oeuvres*, t. 3, pp. 693-785.

³ G. L. Dirichlet, *Zahlentheorie*, ed. 4 (1894), p. 229.

⁴ G. Pall, *Mathematische Zeitschrift*, vol. 36 (1932), p. 321-343.

⁵ Dirichlet, *loc. cit.*

THEOREM 1. *Let m be positive and prime to Δ . The number of representations of m by all the reduced forms of discriminant $-\Delta$ is $\omega \sum_{\mu/m} (-\Delta/\mu)$ where $\omega = 2$, if $\Delta > 4$; $\omega = 4$, if $\Delta = 4$; $\omega = 6$, if $\Delta = 3$, and $(-\Delta/\mu)$ is Kronecker's symbol.*

There are quantities associated with a particular form, invariant in that they are equal for all integers represented by said form which separate the forms of given discriminant into genera which may or may not coincide with the several classes. These invariants, the so-called characters, are defined by

THEOREM 2.⁶ *If p_1, p_2, \dots, p_k are the distinct odd prime factors of Δ , then (n/p_i) has the same value for all integers n prime to Δ , represented by a form $[a, b, c]$ of discriminant $-\Delta$. When Δ is even, $\Delta = -4D$, the same is true of*

$$\begin{aligned} \delta &= (-1)^{1/2(n-1)}, \text{ if } D \equiv 0 \text{ or } 3 \pmod{4} \\ \epsilon &= (-1)^{1/8(n^2-1)}, \text{ if } D \equiv 0 \text{ or } 2 \pmod{8} \\ \delta\epsilon, &\quad \text{ if } D \equiv 0 \text{ or } 6 \pmod{8}. \end{aligned}$$

The set, C_1, C_2, \dots, C_h , will stand for the characters belonging to a certain discriminant, excluding $\delta\epsilon$ if both δ and ϵ are characters. The number of these, h , will equal $k, k+1$, or $k+2$ according to the nature of the discriminant as indicated above. The notation, $C_i(n)$, is to represent the value of the character C_i for n of the form representing n .

All forms of a given discriminant whose characters have the same value are said to form a genus. Since equivalent forms represent the same numbers, all forms in the same class are in the same genus.

When there is a single class in each genus we may proceed using these characters to give the explicit form for the number of representations function for integers m prime to 2Δ . If $[a, b, c]$ represents some integer s , Theorem 2 states $C_i(m) = C_i(s)$ as a necessary condition that m be represented at all by $[a, b, c]$. Since we assume a single class in each genus, each reduced form has different values for the characters. Hence by Theorem 1 we obtain

THEOREM 3. *Let $[a, b, c]$ be a form of discriminant $-\Delta < -4$ such that there is a single class of forms in each genus. If m is an integer prime to 2Δ*

$$\begin{aligned} N[m = ax^2 + bxy + cy^2] \\ = \frac{1}{2^{h-1}} \prod_{i=1}^h [1 + C_i(a)C_i(m)] \sum_{\mu/m} (-\Delta/\mu). \end{aligned}$$

⁶ L. E. Dickson, *Introduction to the Theory of Numbers*, pp. 82, 87.

In order to extend this result to the number of representations of numbers not prime to the discriminant, there are required three auxiliary theorems.

LEMMA 1. Let $[a, b, c]$ be a form of discriminant $-\Delta$. Let p be a prime such that either p^2 divides Δ and $p > 2$ or $p = 2$ and $\Delta \equiv 0$ or $12 \pmod{16}$. Then

$$N[pm = ax^2 + bxy + cy^2, (m, p) = 1] = 0$$

and

$$\begin{aligned} N[p^2m = ax^2 + bxy + cy^2] \\ = N[m = a'x^2 + b'xy + c'y^2] \end{aligned}$$

where $[a', b', c']$ is a form of discriminant $-\Delta/p^2$ whose characters are respectively equal to the corresponding ones for $[a, b, c]$.

LEMMA 2. Let $[a, b, c]$ be a form of discriminant $-\Delta$, and let the prime, p , divide Δ but not satisfy the conditions of Lemma 1. Then

$$\begin{aligned} N[pm = ax^2 + bxy + cy^2] \\ = N[m = a'x^2 + b'xy + c'y^2] \end{aligned}$$

where $[a', b', c']$ is a form of discriminant $-\Delta$ whose characters are equal to the product of the corresponding characters for $[a, b, c]$ and those for the form of discriminant $-\Delta$ representing p .

Lemmas 1 and 2 are taken directly from theorems stated by G. Pall⁷ with our added condition that there be a single class to a genus.

If $[a, b, c]$ has a discriminant $-\Delta \equiv -3 \pmod{8}$, then since $\Delta = b^2 - 4ac$, a, b , and c must be odd, so that if $ax^2 + bxy + cy^2 \equiv 0 \pmod{2}$, then $x^2 + xy + y^2 \equiv 0 \pmod{2}$. If x were odd, y could be neither odd nor even, thus x and y must both be even, $x = 2, y = 2$, and

$$ax^2 + bxy + cy^2 \equiv 0 \pmod{4}.$$

This proves

LEMMA 3. Let $[a, b, c]$ be a form of discriminant $-\Delta \equiv -3 \pmod{8}$. Then

$$\begin{aligned} N[2^r m = ax^2 + bxy + cy^2, m \text{ odd}] \\ = N[m = ax^2 + bxy + cy^2], r \text{ even} \\ = 0, r \text{ odd.} \end{aligned}$$

The number of representations function. In the theorems below giving the explicit form of the number of representations function for all cases where there is a single class of forms to a genus $C_4(s)$ is to stand for the value of

⁷ G. Pall, *Mathematische Zeitschrift*, loc. cit., pp. 331-332, Theorems 4 and 5.

the characters for the positive, primitive form $[a, b, c]$ of discriminant $-\Delta$, and we write $F(m) = \sum_{\mu|m} (-\Delta/\mu)$, where the summation is taken over all divisors μ of m .

THEOREM 4. *If $\Delta = p_1 \cdot p_2 \cdots p_k \equiv 3 \pmod{8}$, $\Delta > 3$, where the p_i are distinct odd primes,*

$$N[2^{2\lambda} p_1^{a_1} \cdots p_k^{a_k} m = ax^2 + bxy + cy^2, (m, 2\Delta) = 1] \\ = \frac{1}{2^{h-1}} \prod_{i=1}^h \{1 + \prod_{j=1}^k [C_i(p_j)]^{a_j} C_i(s) C_i(m)\} F(m).$$

The powers of the odd primes, p_i , in the number represented are reduced according to Lemma 2. Whence by Theorem 3 the statement follows. The even power 2λ is required by Lemma 3. The factor multiplying $F(m)$ occurs in this manner merely to associate the plus or minus one value of the characters with the representation or non-representation according to Lemma 2.

It has been shown previously⁸ that if $\Delta \equiv 7 \pmod{8}$, the only discriminants for which there is a single class to a genus are $\Delta = 7$ and $\Delta = 15$. These are included in

THEOREM 5. *If $\Delta = p_1 p_2$, $p_1 = 3$ or 7 , $p_2 = 5$ or 1 , respectively,*

$$N[2^\lambda p_1^{a_1} p_2^{a_2} m = ax^2 + bxy + cy^2, (m, 2\Delta) = 1] \\ = \frac{\lambda + 1}{2^{h-1}} \prod_{i=1}^h \{1 + [C_i(p_1)]^{a_1} [C_i(p_2)]^{a_2} C_i(s) C_i(m)\} F(m).$$

The powers of the odd primes, p_i , in the number represented are reduced according to Lemma 2, while the power of two is reduced by use of results stated by Dickson⁹ on forms of discriminant -7 and -15 . The theorem follows from Theorem 3.

The only odd discriminants containing the square of a prime as a factor which have a single class to a genus are:

$$\Delta = 27, 75, 99, 147, 315.^{10}$$

These are included in

THEOREM 6. *If $\Delta = p_1^2 p_2 p_3$, where $p_1 = 3, 5, 7, 3$; $p_2 p_3 = 3, 3, 11, 3, 35$; respectively,*

⁸ N. A. Hall, *Mathematische Zeitschrift*, vol. 44 (1938), p. 88.

⁹ Dickson, *loc. cit.*, pp. 81, 88.

¹⁰ Hall, *loc. cit.*

$$\begin{aligned}
 N[2^{2\lambda} p_1^{a_1} p_2^{a_2} p_3^{a_3} &= ax^2 + bxy + cy^2, (m, 2\Delta) = 1] \\
 &= \frac{1}{2^{h-1}} \prod_{i=1}^h \{1 + \prod_{j=1}^3 [C_i(p_j)]^{a_j} C_i(s) C_i(m)\} F(m), \quad \alpha_1 = 0 \\
 &= 0, \quad \alpha_1 = 1 \\
 &= \frac{\omega}{2^{h-2}} \prod_{i=2}^h \{1 + \prod_{j=2}^3 [C_i(p_j)]^{a_j} C_i(s) C_i(p_1^{a_1-2} m)\} F(p_1^{a_1-2} m), \quad \alpha_1 \geq 2
 \end{aligned}$$

where $\omega = 3$ when $p_2 p_3 = 3$ and $\omega = 1$ otherwise, and C_1 is the character associated with p_1 .

The power of p_1 in the number represented is reduced according to Lemma 1, while those of p_2 and p_3 are reduced according to Lemma 2. The statement then follows from Theorem 3. The even power 2λ is required by Lemma 3.

THEOREM 7. *If $\Delta = p_1^\theta p_2 \cdots p_k \equiv 4$ or $8 \pmod{16}$, $\Delta \neq 4$, where $p_1 = 2$, $\theta = 2$ or 3 and the remaining p_i are distinct odd primes,*

$$\begin{aligned}
 N[p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k} m &= ax^2 + bxy + cy^2, (m, \Delta) = 1] \\
 &= \frac{1}{2^{h-1}} \prod_{i=1}^h \{1 + \prod_{j=1}^k [C_i(p_j)]^{a_j} C_i(s) C_i(m)\} F(m).
 \end{aligned}$$

The powers of the primes, p_i , in the number represented are reduced according to Lemma 2. Whence by Theorem 3 the statement follows.

When $\Delta \equiv 0 \pmod{16}$, there is more than a single class to a genus unless $\Delta = 16n$ or $64n$, $n = 1, 3, 7, 15$, or unless $\Delta \equiv 32 \pmod{64}$.¹¹ The latter case is included in

THEOREM 8. *If $\Delta = p_1^5 p_2 \cdots p_k$, where $p_1 = 2$ and the remaining p_i are distinct odd primes,*

$$\begin{aligned}
 N[p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k} m &= ax^2 + bxy + cy^2, (m, \Delta) = 1] \\
 &= \frac{1}{2^{h-1}} \prod_{i=1}^h \{1 + \prod_{j=1}^k [C_i(p_j)]^{a_j} C_i(s) C_i(m)\} F(m), \quad \alpha_1 = 0 \\
 &= 0, \quad \alpha_1 = 1 \\
 &= \frac{1}{2^{h-2}} \prod_{i=1}^{h-2} \{1 + \prod_{j=1}^k [C_i(p_j)]^{a_j} C_i(s) C_i(m)\} F(m), \quad \alpha_1 \geq 2
 \end{aligned}$$

where C_h is the character for forms of discriminant $-\Delta$ not a character for forms of discriminant $-\Delta/4$.

The power of $p_1 = 2$ in the number represented is reduced according to Lemma 1 and those of the odd primes, p_i , according to Lemma 2. The statement is then a consequence of Theorem 3.

When $\Delta \equiv 12 \pmod{16}$ the only discriminants for which there is a single

¹¹ Hall, *loc. cit.*

class of forms to a genus are $\Delta = 12, 28, 60$.¹² These together with $\Delta = 16n$ and $64n$, $n = 3, 7, 15$, and $\Delta = 3$ are included in

THEOREM 9. If $\Delta = 2^\theta p_1 p_2$, where $\theta = 0, 2, 4, 6$, and $p_1 p_2 = 3, 7, 15$,

$$\begin{aligned} N[2^\lambda p_1^{a_1} p_2^{a_2} m = ax^2 + bxy + cy^2, (m, 2\Delta) = 1] \\ &= \frac{m}{2^{h-1}} \prod_{i=1}^h \{1 + \prod_{j=1}^h [C_i(p_j)]^{a_j} C_i(s) C_i(m)\} F(m), \quad \lambda \geq \theta \\ &= \frac{1}{2^{h-1}} \prod_{i=1}^h \{1 + \prod_{j=1}^h [C_i(p_j)]^{a_j} C_i(s) C_i(m)\} F(m), \quad \lambda = \theta - 2 \\ &= \frac{1}{2^h} \prod_{i=1}^{h+1} \{1 + \prod_{j=1}^h [C_i(p_j)]^{a_j} C_i(s) C_i(m)\} F(m), \quad \lambda = \theta - 4 \\ &= \frac{1}{2^{h+1}} \prod_{i=1}^{h+2} \{1 + \prod_{j=1}^h [C_i(p_j)]^{a_j} C_i(s) C_i(m)\} F(m), \quad \lambda = \theta - 6 \\ &= 0, \quad \lambda < \theta, \quad \lambda \text{ odd}, \end{aligned}$$

where

$$\begin{aligned} \omega &= 3, \quad p_1 p_2 = 3, \quad \lambda \text{ even} \\ \omega &= 0, \quad p_1 p_2 = 3, \quad \lambda \text{ odd} \\ \omega &= \lambda - \theta + 1, \quad p_1 p_2 = 7 \text{ or } 15 \end{aligned}$$

and C_1, \dots, C_h are in this case the characters for $\Delta = p_1 p_2$; C_{h+1} and C_{h+2} the additional characters for $\Delta = 16p_1 p_2$ and $64p_1 p_2$ respectively.

The power of 2 in the number represented is reduced according to Lemma 1, and those of p_1 and p_2 according to Lemma 2. These reductions together with the appropriate uses of Theorems 1, 3, and 5 provide the statement given.

The even discriminants $\Delta = 4, 16, 64$ are included in

THEOREM 10. If $\Delta = 4$,

$$\begin{aligned} N[2^a m = ax^2 + bxy + cy^2, (m, 2) = 1] \\ = 4F(m). \end{aligned}$$

If $\Delta = 16$,

$$\begin{aligned} N[2^a m = ax^2 + bxy + cy^2, (m, 2) = 1] \\ = \omega F(m), \end{aligned}$$

$$\omega = 2, \alpha = 0; \quad \omega = 0, \alpha = 1; \quad \omega = 4, \alpha \geq 2.$$

If $\Delta = 64$,

$$\begin{aligned} N[2^a m = ax^2 + bxy + cy^2, (m, 2) = 1] \\ = \frac{1}{2} [1 + \delta(s)\delta(m)] [1 + \epsilon(s)\epsilon(m)] F(m), \quad \alpha = 0 \\ = \omega F(m) \\ \omega = 0, \alpha = 1, 3; \quad \omega = 2, \alpha = 2; \quad \omega = 4, \alpha \geq 4. \end{aligned}$$

The power of 2 in the number represented is reduced according to Lemma 1. Theorem 3 completes the statement.

¹² Hall, *loc. cit.*

The only even discriminants containing as a factor the square of an odd prime which have a single class to a genus are:

$$\Delta = 36, 72, 100, 180, 288.^{13}$$

The number of representations function for these even discriminants is given by

THEOREM 11. If $\Delta = 2^\theta 3^2$, $\theta = 3$ or 5 ,

$$\begin{aligned} N[2^\alpha 3^\beta m = ax^2 + bxy + cy^2, (m, \Delta) = 1] \\ &= \frac{1}{2^{h-1}} \prod_{i=1}^h [1 + C_i(2)^\alpha C_i(s) C_i(m)] F(m), \quad \beta = 0 \\ &= \frac{1}{2^{h-2}} \prod_{i=1}^{h-1} [1 + C_i(2)^\alpha C_i(s) C_i(3^{\beta-2} m)] F(3^{\beta-2} m), \quad \beta \geq 2 \\ &= 0, \quad \beta = 1, \text{ or } \alpha = \theta - 4 \end{aligned}$$

where $C_h = (n/3)$. If $\Delta = 2^2 p^2$, $p = 3$ or 5 ,

$$\begin{aligned} N[2^\alpha p^\beta m = ax^2 + bxy + cy^2, (m, \Delta) = 1] \\ &= \frac{1}{2} \prod_{i=1}^2 [1 + C_i(2)^\alpha C_i(s) C_i(m)] F(m), \quad \beta = 0 \\ &= 4F(p^{\beta-2} m), \quad \beta \geq 2 \\ &= 0, \quad \beta = 1. \end{aligned}$$

If $\Delta = 2^2 \cdot 3^2 \cdot 5$,

$$\begin{aligned} N[2^\alpha 3^\beta 5^\gamma m = ax^2 + bxy + cy^2, (m, \Delta) = 1] \\ &= \frac{1}{2^{h-1}} \prod_{i=1}^h [1 + C_i(2)^\alpha C_i(5)^\gamma C_i(s) C_i(m)] F(m), \quad \beta = 0 \\ &= \frac{1}{2^{h-2}} \prod_{i=1}^{h-1} [1 + C_i(2)^\alpha C_i(5)^\gamma C_i(s) C_i(3^{\beta-2} m)] F(3^{\beta-2} m), \quad \beta \geq 2 \\ &= 0, \quad \beta = 1 \end{aligned}$$

where $C_h = (n/3)$.

The reductions are again made according to Lemma 1 and 2, and the statements follow from Theorem 3.

The numerical calculations for the number of representations will be aided by

THEOREM 12. When there is a single class of forms to a genus and Δ is not divisible by the square of an odd prime, 2^4 , or 2^6 ,

$$F(m) = \sum_{\mu|m} (-\Delta/\mu) = \sum_{\mu|m} \prod_{i=1}^h C_i(\mu).$$

¹³ Hall, *loc. cit.*

When p^2/Δ , p an odd prime,

$$(-\Delta/\mu) = \prod_{i=1}^{h-1} C_i(\mu)$$

where $C_h = (n/p)$.

This theorem is directly evident from the definition of the characters given in Theorem 2 and from the law of quadratic reciprocity.

Numerical computations. As indicated above Theorems 4 through 11 give the explicit form of the number of representations function for all forms of discriminant with a single class to a genus. In specific cases the application of these results will require the knowledge of the characters for the form and the values these characters assume for forms representing various numbers and for different genera of the same discriminant. This information can be readily calculated from the definitions given in Theorem 2. The author has prepared a table giving this data for all known discriminants having a single class to a genus.¹⁴ This table lists all the reduced forms of given discriminant together with the several characters and the values they assume for the numbers represented by each of the reduced forms.

As an illustration of the method consider the form, $2x^2 + 35y^2$, of discriminant $-280 = -2^3 \cdot 5 \cdot 7$. The description of the characters as calculated or read from the table referred to above can be presented compactly:

$280 = 2^3 \cdot 5 \cdot 7$	$\epsilon(n)$	$(n/5)$	$(n/7)$
$[1, 0, 70]$	1	1	1
$[2, 0, 35], 2$	-1	-1	1
$[5, 0, 14], 5$	-1	1	-1
$[7, 0, 10], 7$	1	-1	-1

The reduced forms of discriminant -280 are: $[1, 0, 70]$, $[2, 0, 35]$, $[5, 0, 14]$, $[7, 0, 10]$. The prime factors of the discriminant, 2, 5, and 7, are represented by the last three of these respectively. There are three characters, $\epsilon(n)$, $(n/5)$, $(n/7)$, which take on the values listed for numbers represented by the form opposite. The number of representations function is given for this case by Theorem 7. The function is accordingly:

$$N[2^{a_1}5^{a_2}7^{a_3}m = 2x^2 + 35y^2, (m, 70) = 1] \\ = \frac{1}{4}[1 - (-1)^{a_1+a_2}\epsilon(m)][1 - (-1)^{a_1+a_2}(n/5)][1 + (-1)^{a_2+a_3}(m/7)] \sum_{\mu/m} (-280/\mu)$$

We have, furthermore,

¹⁴ N. A. Hall, *California Institute of Technology, Thesis* (1938), pp. 104-116.

$$\begin{aligned}
 \epsilon(m) &= 1, & m &\equiv 1 \text{ or } 7 \pmod{8} \\
 &= -1, & m &\equiv 3 \text{ or } 5 \pmod{8} \\
 (m/5) &= 1, & m &\equiv 1 \text{ or } 4 \pmod{5} \\
 &= -1, & m &\equiv 2 \text{ or } 3 \pmod{5} \\
 (m/7) &= 1, & m &\equiv 1, 2, \text{ or } 4 \pmod{7} \\
 &= -1, & m &\equiv 3, 5, \text{ or } 6 \pmod{7}.
 \end{aligned}$$

Hence we may separate integers, m , $(m, 70)$, into residue classes, modulo 280, with the triplet $\epsilon(m)$, $(m/5)$, $(m/7)$, identical in value for all integers in the class:

	$\epsilon(m)$	$(m/5)$	$(m/7)$	
1)	+	+	+	1, 9, 39, 71, 79, 81, 121, 151, 169, 191, 239, 249.
2)	—	+	+	11, 29, 51, 99, 109, 141, 149, 179, 211, 219, 261, 221.
3)	+	—	+	23, 57, 113, 127, 177, 183, 193, 207, 233, 247, 263, 137.
4)	—	—	+	37, 53, 67, 93, 107, 123, 163, 197, 253, 267, 277, 43.
5)	+	+	—	31, 41, 111, 89, 129, 159, 199, 201, 209, 241, 271, 279.
6)	—	+	—	19, 59, 61, 69, 101, 131, 139, 171, 181, 229, 251, 269.
7)	+	—	—	17, 33, 47, 73, 87, 97, 103, 153, 223, 257, 143, 167.
8)	—	—	—	3, 13, 27, 83, 117, 157, 173, 187, 213, 227, 237, 243.

According to the parity of $\alpha_1, \alpha_2, \alpha_3$ we have the four cases:

	$\alpha_1 + \alpha_2$	$\alpha_1 + \alpha_3$	$\alpha_2 + \alpha_3$	α_1	α_2	α_3
a)	even	even	even	even	even	even
				odd	odd	odd
b)	odd	odd	even	odd	even	even
				even	odd	odd
c)	odd	even	odd	even	odd	even
				odd	even	odd
d)	even	odd	odd	odd	odd	even
				even	even	odd

Applying these results to the number of representations function as given above, it is possible to state further:

$$\begin{aligned} N[2^{\alpha_1} 5^{\alpha_2} 7^{\alpha_3} m] &= 2x^2 + 35y^2, (m, 70) = 1 \\ &= 2 \sum_{\mu|m} (-280/\mu) \end{aligned}$$

when $\alpha_1, \alpha_2, \alpha_3$ and m are paired according to the divisions above: a), 4); b), 1); c), 7); d), 6); otherwise the number of representations is zero.

According to Theorem 12,

$$F(m) = \sum_{\mu|m} \epsilon(\mu) (\mu/5) (\mu/7) = \sum_{\mu|m} (-280/\mu).$$

Hence $F(m)$ is equal to the excess of the divisors of m in classes 1), 4), 6), 7) over those in classes 2), 3), 5), 8).

The formula may be illustrated further by verifying the number of representations of 23902. If

$$23902 = 2x^2 + 35y^2,$$

the empirically obtained solutions are: $x = \pm 11, y = \pm 26$; $x = \pm 59, y = \pm 22$; $x = \pm 101, y = \pm 10$; $x = \pm 109, y = \pm 2$; with all choices of sign permissible. Hence the number of representations is 16. To check this with our formula, we observe that $23902 = 2 \cdot 17 \cdot 19 \cdot 37$; hence

$$\begin{aligned} N[23902] &= 2x^2 + 35y^2 \\ &= N[2 \cdot 17 \cdot 19 \cdot 37] = 2x^2 + 35y^2 \\ &= 2 \sum_{\mu|11951} \epsilon(\mu) (\mu/5) (\mu/7). \end{aligned}$$

Since, referring to the previous notation, $17 \cdot 19 \cdot 37 = 11951 \equiv 191 \pmod{280}$, we have case b), 1). Furthermore, for each of the prime factors of 11951, $\epsilon(\mu) (\mu/5) (\mu/7)$ is seen by reference to 4), 6), and 7) to be +1. Hence the same is true for all factors. In all, 11951 has eight factors: 1; 17; 19; 37; $17 \cdot 19$; $17 \cdot 37$; $19 \cdot 37$; $17 \cdot 19 \cdot 37$. Thus, finally

$$N[23902] = 2x^2 + 35y^2 = 2 \cdot 8 = 16,$$

to agree with the empirical result.

QUEENS COLLEGE,
FLUSHING, NEW YORK.

PARTITION HYPERGROUPS.*¹

By HOWARD CAMPAIGNE.

1. Introduction. In 1934 F. Marty² and H. S. Wall³ introduced independently the notion of hypergroup. They both used this term for a system which may in particular be a group, but in which the product of two elements is in general a set of elements of the system. Both writers discussed partition hypergroups in which the elements are sets of elements of a group. The question arises as to whether or not every hypergroup can be represented in this way as a partition hypergroup obtained from a group. Marty offered the conjecture that the answer is in the affirmative. Wall gave an example of a hypergroup which cannot be so represented by means of the *special* conjugation which he considered. *It is shown in the present paper (section 6) that it is not possible to represent this hypergroup by means of any conjugation whatsoever among the elements of a group.*

The second main result obtained is a characterization of simple groups in terms of a partition hypergroup (section 9). It is shown that a group G is simple if and only if a certain partition hypergroup contains no proper sub-hypergroups except the identity group. This partition hypergroup is obtained by means of a conjugation among the elements of G depending on its group of inner automorphisms, and the proof depends on the study of the lattices of this hypergroup.

Sections 3, 4, 5 treat of partition hypergroups, the mapping of one hypergroup upon another, semi-regular, regular, and commutative hypergroups, respectively. In section 7 there are considered examples of conjugations.

An analogue of the direct product of groups is the subject of section 8. Hypergroups which are products of two hypergroups are completely characterized. Many of the ideas of this section are generalizations of ideas in R. Remak's papers (there cited).

* Received August 4, 1938; Revised February 21, 1940.

¹ Presented to the Society April 8, 1938.

² F. Marty, "Sur une généralisation de la notion de groupe," *Särtryck ur Föreläsningar vid Åttionde Skandinaviska Matematikerkongressen i Stockholm* (1934), pp. 45-49.

³ H. S. Wall, "Hypergroups," *Bulletin of the American Mathematical Society*, vol. 41 (1935), p. 36. [Presented at the annual meeting of the American Mathematical Society, Pittsburgh, December 27-31, 1934.]

2. Definition of a hypergroup. We consider a system H of elements a, b, c, \dots in which a product ab is defined for every pair of elements a, b of H . The product CD of two subsets C, D of H is defined as the set of all distinct elements of the products cd as c ranges over C and d over D . The system H is a *hypergroup* if it satisfies the following postulates.⁴

I. If a and b are elements of H , then the product ab is a non-vacuous subset of distinct elements of H .

II. If a, b, c are elements of H then $a(bc) = (ab)c$.

III. There is in H at least one element e , called an identity, such that for every element b in H the products eb and be contain b .

IV. There is at least one identity e in H such that if b is an arbitrary element of H there is in H at least one element b^{-1} , called an inverse of b relative to e , such that the sets $b^{-1}b$ and bb^{-1} contain e .

It is easily shown that if ⁵ $a_e H, b_e H$ then there exist elements x, y in H such that $b_e ax, b_e ya$. From the definition of a group it follows that if the products ab are all single element sets, then H is a group.

A *subhypergroup* K of H is a subset of H in which the postulates I to IV are satisfied with the law of multiplication of H . If $K \neq H$ and contains at least one identity of H , K is a *proper* subhypergroup of H .

3. Definition of conjugation. An equivalence relation, \sim , in a hypergroup H is called a *conjugation* if when $a_e H, b_e H, c_e ab, c' \sim c$, then there exist elements a', b' in H such that $a' \sim a, b' \sim b, c'_e a'b'$. If $a' \sim a$ we shall say that a' is *conjugate* to a . This relation is symmetric, reflexive, and transitive.

If γ is a conjugation in H , the distinct residue classes $\{a\}_\gamma$ of elements conjugate to a form a hypergroup $\{H\}_\gamma$, with respect to the law of multiplication which requires that

$$\{c\}_\gamma \{a\}_\gamma \{b\}_\gamma$$

if and only if there exist elements a', b' in H conjugate to a and b , respectively, such that $c'_e a'b'$. It will be seen that postulates I to IV are satisfied by this system. If e is an identity of H then $\{e\}_\gamma$ is obviously an identity of $\{H\}_\gamma$.

⁴ This definition is somewhat different both from that of Wall and from that of Marty. Wall's definition ("Hypergroups," *American Journal of Mathematics*, vol. 59 (1937), pp. 77-98) differs only in one respect, namely, that he requires the product ab to have exactly n elements (not necessarily distinct) where n is a fixed integer greater than 0. The definition we have adopted agrees with that of the regular multigroup of Dresher and Ore, "Theory of multigroups," *American Journal of Mathematics*, vol. 60 (1938), pp. 705-733.

⁵ The symbol $a_e A$ is read " a is an element of A ."

if b^{-1} is an inverse of b relative to e then $\{b^{-1}\}_\gamma$ is an inverse of $\{b\}_\gamma$ relative to $\{e\}_\gamma$. We shall call $\{H\}_\gamma$ the *partition hypergroup of H relative to the conjugation γ* .

By the remark near the end of section 2, $\{H\}_\gamma$ is a group if and only if when a, b, a', b' are elements of H such that $a \sim a', b \sim b', c_e ab, d_e a' b'$, then $c \sim d$.

We shall say that a subset J of a hypergroup H is appropriate relative to a conjugation γ in H if J contains with a all the conjugates of a relative to γ . If J is an appropriate subhypergroup of H then γ induces a conjugation γ_1 in J , and it is easily seen that the partition hypergroup $\{J\}_{\gamma_1}$ is a subhypergroup of $\{H\}_\gamma$. Conversely, if K is a subhypergroup of $\{H\}_\gamma$ then the set J of all the elements of H contained in the residue classes of K is appropriate relative to γ .

4. Mapping of one hypergroup upon another. We shall consider a mapping of a hypergroup A upon a hypergroup \mathfrak{A} such that the following conditions are satisfied.

(1) Each element a of A is mapped upon a uniquely determined element α of \mathfrak{A} , in symbols $a \rightarrow \alpha$.

(2) If $\alpha_e \mathfrak{A}$ then there is at least one element a of A such that $a \rightarrow \alpha$.

(3) If $c_e ab$ and $c \rightarrow \alpha, a \rightarrow \alpha, b \rightarrow \beta$ then $\alpha_e \alpha \beta$.

(4) If $\alpha_e \mathfrak{A}, \beta_e \mathfrak{A}$ and $\alpha_e \alpha \beta$ then there exist elements a, b, c in A such that $a \rightarrow \alpha, b \rightarrow \beta, c \rightarrow \alpha$ and $c_e ab$.

It follows from (2), (3) that an identity of A is mapped upon an identity of \mathfrak{A} . If a^{-1} is an inverse of a relative to an identity e , and $a^{-1} \rightarrow \alpha_1, a \rightarrow \alpha, e \rightarrow e$, then it follows from (2), (3) that $\alpha_1 = \alpha^{-1}$ is an inverse of α relative to e .

If there exists a mapping of A upon \mathfrak{A} satisfying conditions (1) to (4) we shall say that A is *semi-isomorphic* with \mathfrak{A} , in symbols $A \cong \mathfrak{A}$. In particular A is *isomorphic* with \mathfrak{A} , $A \approx \mathfrak{A}$, if the mapping satisfies (1), (3), and (4), and the following condition stronger than (2):

(2') If $\alpha_e \mathfrak{A}$ then there is exactly one element a of A such that $a \rightarrow \alpha$.

Isomorphism is reflexive, symmetric, and transitive. Semi-isomorphism is reflexive but not symmetric. It is easily seen to be transitive. In fact, if $P \cong Q, Q \cong R, p \rightarrow q, q \rightarrow r$, then it will be seen that the mapping $p \rightarrow r$ maps P upon R in such a way that the conditions (1) to (4) are satisfied, and therefore $P \cong R$.

THEOREM 4.1. *If A and \mathfrak{A} are finite hypergroups, and $A \cong \mathfrak{A}$, $\mathfrak{A} \cong A$, then $A \approx \mathfrak{A}$.*

Proof. Let A, \mathfrak{A} be of orders μ, ν respectively. Since $A \cong \mathfrak{A}$ it follows that $\mu \geq \nu$, and since $\mathfrak{A} \cong A$, $\mu \leq \nu$. Hence $\mu = \nu$, and therefore $A \approx \mathfrak{A}$ since the mapping must be one to one.

The following theorem may readily be verified.

THEOREM 4.2. *Let $\{H\}_\gamma$ be a partition hypergroup of H relative to the conjugation γ . If $a \in H$, then the mapping $a \rightarrow \{a\}_\gamma$ is a semi-isomorphism, so that $H \cong \{H\}_\gamma$.*

THEOREM 4.3. *Let $\{H\}_{\gamma_1}, \{H\}_{\gamma_2}$ be partition hypergroups of H relative to conjugations γ_1, γ_2 such that if $a \sim b$ relative to γ_1 then $a \sim b$ relative to γ_2 . Then there exists a conjugation γ_3 in $\{H\}_{\gamma_1}$ such that $\{\{H\}_{\gamma_1}\}_{\gamma_3} \approx \{H\}_{\gamma_2}$.*

Proof. Let $\{a\}_{\gamma_1} \sim \{b\}_{\gamma_1}$ when $a \sim b$ relative to γ_2 . This defines a conjugation γ_3 in $\{H\}_{\gamma_1}$. The mapping $\{\{a\}_{\gamma_1}\}_{\gamma_3} \rightarrow \{a\}_{\gamma_2}$ of $\{\{H\}_{\gamma_1}\}_{\gamma_3}$ upon $\{H\}_{\gamma_2}$ is seen to be an isomorphism.

An automorphism of H is an isomorphic mapping of H upon itself. The set of automorphisms of H forms a group. There is one case in which a subgroup P of this group induces a group of automorphisms in a partition hypergroup $\{H\}_\gamma$, namely, when the conjugation is preserved under the automorphisms of P . If an automorphism p of P maps a upon a' we shall write $a \xrightarrow{p} a'$. Suppose then that whenever $p \in P$, $a \xrightarrow{p} a'$, $b \xrightarrow{p} b'$, $a \sim b$ it follows that $a' \sim b'$. Define a mapping p' of $\{H\}_\gamma$ upon itself by letting $\{a\}_\gamma \xrightarrow{p'} \{a'\}_\gamma$ when $a \xrightarrow{p} a'$. This is clearly a one to one mapping of $\{H\}_\gamma$ upon itself, and is easily seen to be an automorphism of $\{H\}_\gamma$. The set of these induced automorphisms forms a group Q . It may be shown that $P \approx Q$ if when $p \in P$, $a \in H$, $a \xrightarrow{p} b$ then $a \not\sim b$.

5. Semi-regular, regular, and commutative hypergroups. A hypergroup A will be called *semi-regular* if it contains at least one element s , called a *scalar*, such that if $a \in A$ then as and sa are single element sets. The set of all scalars is called the *nucleus* of A . Wall⁶ has shown that for his hypergroup the nucleus forms a subgroup of the hypergroup, and its identity is the only identity of the hypergroup. The same holds for the hypergroup here considered, and the proof given by Wall holds without modification.

THEOREM 5.1. *Let A, \mathfrak{A} be two hypergroups such that there is a semi-isomorphic mapping $a \rightarrow \alpha$ of A upon \mathfrak{A} . Let E, B denote the subsets of elements of A mapped upon an identity e and an arbitrary element b , respectively, of \mathfrak{A} . Then \mathfrak{A} is semi-regular if and only if*

⁶ Wall, *loc. cit.*, in footnote 4, Theorem 4, p. 79.

- (1) E is a subhypergroup of A , and
- (2) $EB \subset B$, $BE \subset B$ for every b .

Proof. Supposing that \mathfrak{H} is semi-regular, we shall prove that (1) and (2) hold. Let a, b be elements of E , and $c \in ab$. Then $a \rightarrow e$, $b \rightarrow e$ and therefore $c \rightarrow e = e^2$, that is, c is in E . If $a \in E$ and a^{-1} an inverse of a relative to an identity e (necessarily in E), then if $a^{-1} \rightarrow a_1$ we have: $e \in aa^{-1}$, $e \in ea_1 = a_1 = e$, so that $a^{-1} \in E$. This completes the proof of (1). To prove $EB \subset B$, let $b \rightarrow b$ for every element b of B . Then if $a \in E$ and $c \in ab$ we must have $c \in ab = eb = b$ so that $c \in B$. Similarly, $BE \subset B$.

Conversely let (1) and (2) hold. To prove that \mathfrak{H} is semi-regular we shall show that e is a scalar. If $c \in eb$ then, since $A \cong \mathfrak{H}$, there exist elements c, a, b in A such that $c \rightarrow c$, $a \in E$, $b \in B$, and $c \in ab$. Thus $c \in EB \subset B$ or $c \rightarrow b = c$ and therefore $b = eb$. Similarly, $b = be$. Since this holds for every b in \mathfrak{H} it follows that e is a scalar, as was to be proved.

If $A \cong \mathfrak{H}$ and \mathfrak{H} is semi-regular, then

COROLLARY 5.1. *The set N of all elements of A which are mapped upon elements of the nucleus of \mathfrak{H} is a subhypergroup of A .*

A semi-regular hypergroup is called *regular* if each element has a unique inverse with respect to the identity e and if $e \in ab$ implies $e \in ba$.

THEOREM 5.2. *If H is a hypergroup such that $e \in ab$ implies that $e \in ba$ for every identity e , then a partition hypergroup $\{H\}_\gamma$ is regular if and only if:*

(1) *the identities of H are all contained in a single class $\{e\}_\gamma$, and the set E of elements in this residue class is a subhypergroup of H ;*

(2) *if B is the set of elements in any class $\{b\}_\gamma$, then $EB \subset B$ and $BE \subset B$;*

(3) *the inverses relative to all identities of the elements in any class $\{a\}_\gamma$ are all in one and the same class $\{a^{-1}\}_\gamma$.*

Proof. By Theorem 4.2 the mapping $a \rightarrow \{a\}_\gamma$ is a semi-isomorphism of H upon $\{H\}_\gamma$. Hence by Theorem 5.1 the conditions (1) and (2) are necessary for the regularity of $\{H\}_\gamma$. Condition (3) is also necessary. For if $\{a\}_\gamma \in \{H\}_\gamma$ then when $\{H\}_\gamma$ is regular $\{a^{-1}\}_\gamma$ must be the unique inverse of $\{a\}_\gamma$, and no other class can contain an inverse of an element in the class $\{a\}_\gamma$.

Conditions (1) and (2) are sufficient for the semi-regularity of $\{H\}_\gamma$ by Theorem 5.1. To prove that (3) implies the regularity of $\{H\}_\gamma$ we must show that every element $\{a\}_\gamma$ has a unique inverse. If $\{e\}_\gamma \in \{a\}_\gamma \{b\}_\gamma$ (so that by hypothesis $\{e\}_\gamma \in \{a\}_\gamma \{b\}_\gamma$), then there exist elements a', b' conjugate

to a and b such that $e_a a' b'$. Hence b' is the inverse of a' , and therefore by (3), $\{b'\}_\gamma = \{a^{-1}\}_\gamma = \{b\}_\gamma$.

A hypergroup is *commutative* if $ab = ba$ for every pair a, b of its elements. It is easy to see that if A is commutative and $A \cong \mathfrak{A}$, then \mathfrak{A} is commutative. In particular, a partition hypergroup of a commutative hypergroup is commutative.

6. A regular commutative hypergroup which cannot be represented as a partition of a group. Such a hypergroup is given by the following table:

	e	b	a
e	e	b	a
b	b	a	e, b
a	a	e, b	a, b

Wall⁷ showed that this hypergroup cannot be represented as a partition hypergroup of a group relative to the special conjugation which he considered. We shall prove that this hypergroup has a property which no partition hypergroup of a group has, namely: it is not inversive.

A hypergroup is *inversive*⁸ if when $c_e ab$ there exists an identity e , and inverses c^{-1}, a^{-1}, b^{-1} , of c, a, b , such that $c^{-1} e b^{-1} a^{-1}$. Every group is necessarily inversive. We shall prove that every partition hypergroup of a group is inversive. More generally, we have

THEOREM 6.1. *If $A \cong \mathfrak{A}$ and A is inversive, then \mathfrak{A} is inversive.*

Proof. If $c_e ab, a_e \mathfrak{A}, b_e \mathfrak{A}$, and $c \rightarrow c, a \rightarrow a, b \rightarrow b, c_e ab$, then by hypothesis there exist inverses a^{-1}, b^{-1}, c^{-1} relative to an identity e such that $c^{-1} e b^{-1} a^{-1}$. Let $c^{-1} \rightarrow c_1, b^{-1} \rightarrow b_1, a^{-1} \rightarrow a_1$. Then $c_1 e b_1 a_1$. Now $e \rightarrow e$, where e is an identity of \mathfrak{A} , and also $a^{-1} \rightarrow a^{-1}, b^{-1} \rightarrow b^{-1}, c^{-1} \rightarrow c^{-1}$, that is, $a_1 = a^{-1}, b_1 = b^{-1}$, and $c_1 = c^{-1}$. Thus $c^{-1} e b^{-1} a^{-1}$, so that \mathfrak{A} is inversive.

In the example above, $a_e a^2$ but $a^{-1} \nmid (a^{-1})^2$, so that the hypergroup is not inversive. We therefore have:

THEOREM 6.2. *There exists a regular commutative hypergroup which is not isomorphic with a partition hypergroup of a group.*

7. Examples of conjugations. Other writers⁹ have discussed a conjugation of which the following is an immediate generalization.

⁷ *Loc. cit.*, p. 96.

⁸ This is less restrictive than Dresher and Ore's *reversible in itself*. See the reference cited in footnote 4, p. 717.

⁹ Wall, *loc. cit.*, pp. 92-93. Marty, "Sur les groupes et hypergroupes attachés à une fraction rationnelle," *Annales de l'Ecole Normale Supérieure* (3), vol. 53 (1936), pp. 83-123. A generalization is mentioned by Dresher and Ore, p. 720.

Example 1. Let H be semi-regular, and S, T subgroups of its nucleus. Let $a \sim b$ if $a = sbt$ where $s \in S, t \in T$. This defines a conjugation in H . Denote the partition hypergroup by $\{H; S, T\}$. In particular, if H is a group, S an invariant subgroup, and T the identity group, then $\{H; S, T\}$ is the quotient group H/S . The partition hypergroup $\{H; S, T\}$ is semi-regular if $S = T$.

Example 2. Let H be commutative and inversive and contain an identity e such that each element has exactly one inverse with respect to e . Let $b \sim a$ if $b = a$ or $b = a^{-1}$. This relation is a conjugation in H , and defines a partition hypergroup $\{H\}$. In this case $H \approx \{H\}$ if and only if every element of H is self-inverse with respect to e . If H is an Abelian group then $\{H\}$ is a regular commutative hypergroup in which the product of any two elements is a set of at most two elements.

Example 3. We may define a conjugation in an arbitrary hypergroup H in terms of any subgroup P of the group of automorphisms of H . Inasmuch as the partition hypergroups obtained in this way play an important role in a subsequent result, we shall develop here some of their properties.

The conjugation is defined as follows. Let $a \sim b$ if a is mapped on b by some automorphism of P . This relation is clearly a conjugation, and so defines a partition hypergroup of H which we shall denote by $\{H\}_P$. By Theorem 4.3 we have at once:

THEOREM 7.1. *If Q is a subgroup of a group P of automorphisms of H , then there exists a conjugation γ in $\{H\}_Q$ such that $\{\{H\}_Q\}_\gamma \approx \{H\}_P$.*

THEOREM 7.2. *If H is semi-regular (regular) then $\{H\}_P$ is semi-regular (regular).*

Proof. The identity e of a semi-regular hypergroup is mapped on itself by every automorphism, and therefore the class $\{e\}_P$ contains only e . Evidently $\{e\}_P$ is a scalar of $\{H\}_P$, so that $\{H\}_P$ is semi-regular.

If H is regular then we must show in addition that every element of $\{H\}_P$ has exactly one inverse, and that $\{e\}_P \{a\}_P \{b\}_P$ implies that $\{e\}_P \{b\}_P \{a\}_P$. If $\{e\}_P \{a\}_P \{b\}_P$ then there exist elements a', b' conjugate to a, b such that $e a' b'$, and hence the class $\{b\}_P$ contains the inverse of an element of $\{a\}_P$. But if a is mapped on a' by an automorphism p_1 , then a^{-1} is mapped on a'^{-1} by p_1 . Since $a'^{-1} = b' \sim b$ it then follows that $a^{-1} \sim b$, that is, $\{b\}_P = \{a^{-1}\}_P$. The regularity of $\{H\}_P$ follows.

THEOREM 7.3. *If G is a group and P a group of its automorphisms, then $\{G\}_P$ is a regular hypergroup, and is a group if and only if P is the identity group.*

Proof. The first part is a corollary to theorem 7.2. To prove the last part, let us suppose $\{G\}_P$ is a group if $\{a\}_P \in \{G\}_P$ then $\{a\}_P \{a^{-1}\}_P = \{e\}_P$. Since $\{e\}_P$ contains but one element, we must have $a'a^{-1} = e$ if $a' \sim a$, so that $a' = a$, and hence P contains only the identity automorphism. The converse is obviously true.

8. Product hypergroups.¹⁰ The product $A \times B$ of two hypergroups is the set of all ordered couples $a \times b$ where $a \in A$, $b \in B$, and where multiplication between couples is defined by agreeing that $a \times b \in (a_1 \times b_1)(a_2 \times b_2)$ if $a \in a_1 a_2$, $b \in b_1 b_2$. It is easily seen that $A \times B$ is a hypergroup. A subhypergroup H of $A \times B$ is called a *sub-product* of A and B if each element of A (and likewise each element of B) is represented in at least one couple $a \times b$ in H . A sub-product of A and B will be denoted by $A \times B$.

We shall begin by listing, without proof, some of the more obvious properties of products and sub-products.

- (1) $A \times B$ is a group if and only if A and B are groups.
- (2) $A \times B \approx B \times A$.
- (3) If A_1 is a subhypergroup of A , then $A_1 \times B$ is a subhypergroup of $A \times B$.
- (4) If K is a subhypergroup of $A \times B$, then there exist subhypergroups A_1 and B_1 of A and B such that K is a sub-product of A_1 and B_1 .
- (5) A sub-product $A \times B$ is semi-regular only if A and B are semi-regular. The nucleus of $A \times B$ is a sub-product of subgroups of the nuclei of A and B . $A \times B$ is semi-regular (regular) if and only if A and B are semi-regular (regular), and its nucleus is $M \times N$, where M and N are the nuclei of A and B .

If B contains an "idempotent" element b such that $b^2 = b$, then it is evident that $A \times B$ contains a subhypergroup isomorphic with A , namely $A \times b$. The converse is not necessarily so, as shown by the following example. Let T_ω be the hypergroup of ω elements $t_0, t_1, \dots, t_{\omega-1}$, where for every λ, μ, ν we have $t_\lambda \in t_\mu t_\nu$. Let T_∞ be a set of a countable number of elements t_ν with multiplication similarly defined. T_ω has no subhypergroups, and no idempotent elements. Yet $T_\infty \times T_\omega \approx T_\infty$ under the correspondence

¹⁰ See Robert Remak's papers, "Über minimale invariante Untergruppen in der Theorie der Endlichen Gruppen," vol. 162 (1930), pp. 1-16, and "Über die Darstellung der endlichen Gruppen als Untergruppen direkter Producte," vol. 163 (1930), pp. 1-44 of the *Journal für Mathematik*.

$t_\mu' \times t_\nu \rightarrow t_{\mu\omega+\nu'}$. Note that the product $T_\mu \times T_\nu \approx T_{\mu\nu}$ has no subhypergroups. The following theorem gives conditions under which the converse is true.

THEOREM 8.1. *Let every descending chain of subhypergroups $A \supset A' \supset A'' \supset \dots$ of A be finite, and let A contain an identity e such that e^2 is a finite set. If $A \times B$ contains a subhypergroup isomorphic with A , then B must contain an idempotent element.*

Proof. Let K_0 be a subhypergroup of $A \times B$ such that $K_0 \approx A$. Then by (4) A, B contain subhypergroups A', B' such that $K_0 = A' \times B'$, a subproduct of A' and B' . If $A' = A$ the argument proceeds as in the next paragraph. If $A' \neq A$, then $K_0 \supset K_1 \approx A'$, since $K_0 \approx A$. Therefore by (4), A', B' contain subhypergroups A'', B'' such that $K_1 = A'' \times B''$. If $A'' \neq A'$, then $K_1 \supset K_2 \approx A''$. Therefore by (4) A'', B'' contain subhypergroups A''', B''' such that $K_2 = A''' \times B'''$. Continuing in this way we get a descending chain of subhypergroups $A \supset A' \supset A'' \supset \dots$ which must terminate. Therefore there is a ν such that $A^{(\nu-1)} = A^{(\nu)}$. Without loss of generality we can assume that $A' = A$.

Let $a \rightarrow a' \times b'$ be corresponding elements under the isomorphism $A \approx A \times B' = K_0$. If a_η is an identity in A then $a_\eta' \times b_\eta'$ is an identity in K_0 , and a_η' and b_η' are identities in A and B' respectively. Let $a_1, a_2, \dots, a_\eta, \dots$ be the identities of A . Let $\alpha_\eta, \beta_\eta, \gamma_\eta$ be the numbers of elements in the sets $a_\eta'^2, b_\eta'^2$, and a_η^2 . Thus $\alpha_\eta, \beta_\eta, \gamma_\eta$ are positive integers (or infinite) and $\alpha_\eta \beta_\eta = \gamma_\eta$. There is a smallest γ_η , let it be γ_λ . Since $\alpha_\lambda \beta_\lambda = \gamma_\lambda$ we have $\alpha_\lambda \leq \gamma_\lambda$. But α_λ is also among the integers γ_η , since a_η' is an identity in A . Thus $\alpha_\lambda \geq \gamma_\lambda$, and $\alpha_\lambda = \gamma_\lambda$, so that $\beta_\lambda = 1$. Since b_λ' is an identity, $b_\lambda'^2 = b_\lambda'$, and B has an idempotent element.

COROLLARY 8.1. *Let $A \times B$ satisfy the descending chain condition, and have an identity e such that e^2 is a finite set. Then $A \times B$ contains subhypergroups A_0 and B_0 , isomorphic with A and B respectively, if and only if $A \times B$ contains an idempotent element, the intersection of A_0 and B_0 .*

We next consider the question: when can a hypergroup be expressed as a product? In order to get an answer to our question we must first consider conjugations in product hypergroups, which can always be expressed in terms of conjugations in the factor hypergroups, according to the following theorem.

THEOREM 8.2. *If A and B are hypergroups with conjugations among their elements, then there is a conjugation among the elements of $A \times B$ such that $\{A\} \times \{B\} \approx \{A \times B\}$. Conversely, if there is a conjugation among*

the elements of $A \times B$ then there exist conjugations among the elements of A and of B such that $\{A \times B\} \approx \{A\} \times \{B\}$.

Proof. Let $a \times b \sim a' \times b'$ if, and only if, $a \sim a'$ and $b \sim b'$. The mapping $\{a \times b\} \rightarrow \{a\} \times \{b\}$ is seen to be an isomorphism.

COROLLARY 8.2. *There exists a conjugation in $A \times B$ such that $\{A \times B\} \approx A$. That is, $A \times B \cong A$.*

Proof. In A let $a \sim a'$ if $a = a'$, and in B let $b \sim b'$ for every b and b' .

The following theorem is a direct generalization from standard group theory.

THEOREM 8.3. *Necessary and sufficient conditions that a semi-regular hypergroup H be the product of hypergroups A and B are:*

- (1) A and B are semi-regular and contained in H ;
- (2) if $a \in A$ and $b \in B$ then $ab = ba$ is a single element;
- (3) $AB = H$, and $a_1 b_1 = a_2 b_2$ only if $a_1 = a_2$ and $b_1 = b_2$.

Proof. To establish the sufficiency of these conditions consider $A \times B$. Each element of H is uniquely representable as a product ab . The mapping $a \times b \rightarrow ab$ is seen to be an isomorphism. The necessity is easily seen.

THEOREM 8.4. *A necessary and sufficient condition that an arbitrary hypergroup H be isomorphic with the product of two hypergroups is that there be two conjugations γ_1 and γ_2 in H with the following properties. For every pair x and y in H there exists a unique element c such that $c \sim x$ relative to γ_1 and $c \sim y$ relative to γ_2 . If $x_3 \in x_1 x_2$, and $y_3 \in y_1 y_2$, then $c_3 \in c_1 c_2$. Then $H \approx \{H\}_{\gamma_1} \times \{H\}_{\gamma_2}$.*

Proof of necessity. In $A \times B$ define γ_1 by $a \times b \sim a' \times b'$ when $a = a'$, and γ_2 by $a \times b \sim a' \times b'$ when $b = b'$. These conjugations satisfy the conditions above, and $\{A \times B\}_{\gamma_1} \approx A$, $\{A \times B\}_{\gamma_2} \approx B$.

Proof of sufficiency. Consider $\{H\}_{\gamma_1} \times \{H\}_{\gamma_2}$. Since the classes $\{x\}_{\gamma_1}$ and $\{y\}_{\gamma_2}$ have just one element c in common, the pair $\{x\}_{\gamma_1} \times \{y\}_{\gamma_2}$ can be represented uniquely as $\{c\}_{\gamma_1} \times \{c\}_{\gamma_2}$. The mapping $\{c\}_{\gamma_1} \times \{c\}_{\gamma_2} \rightarrow c$ is an isomorphism of $\{H\}_{\gamma_1} \times \{H\}_{\gamma_2}$ with H .

We conclude with conditions under which the product is commutative or inversive.

THEOREM 8.5. *A necessary and sufficient condition that $A \times B$ be commutative is that both A and B be commutative.*

THEOREM 8.6. *A necessary and sufficient condition that $A \times B$ be inverse is that both A and B be inverse.*

9. The lattices of a hypergroup. The structure of a hypergroup is clarified by studying its lattice of subsets. A *lattice* is defined as a set \mathfrak{C} of elements A, B, C, \dots such that the following conditions are satisfied.

(1) For each pair of elements A and B in \mathfrak{C} there exist elements $A \vee B$ and $A \wedge B$ in \mathfrak{C} , called respectively the *union* and *intersection* of A and B .

(2) These combinations are commutative, $A \vee B = B \vee A$ and $A \wedge B = B \wedge A$.

(3) They are associative as well, $A \vee (B \vee C) = (A \vee B) \vee C$ and $A \wedge (B \wedge C) = (A \wedge B) \wedge C$.

(4) For each pair A and B we have $A \vee (B \wedge A) = A = (A \vee B) \wedge A$.

The intersection $C \wedge D$ of two subsets C and D of a hypergroup H is the set of all elements common to the two. The union ¹¹ $C \vee D$ is the set of all elements contained in products $x_1 x_2 \dots x_\mu$, μ any integer, where x_η is an element of either C or D . The closed subsets of a hypergroup A form a lattice ¹² \mathfrak{M} .

If \mathfrak{M} and \mathfrak{B} are lattices the set of all pairs of elements of \mathfrak{M} and \mathfrak{B} form a lattice $\mathfrak{M} \times \mathfrak{B}$, their *direct join*.¹³ If \mathfrak{M} is the lattice of the closed subsets of a hypergroup A , and \mathfrak{B} is that of the hypergroup B , what is the relation between $A \times B$ and $\mathfrak{M} \times \mathfrak{B}$? To answer this we define a plenary subset of $A \times B$ as one which is the product $H \times J$ of its component sets. If $H_1 \times J_1$ and $H_2 \times J_2$ are two plenary subsets of $A \times B$ then

$$\begin{aligned}(H_1 \times J_1) \vee (H_2 \times J_2) &= (H_1 \vee H_2) \times (J_1 \vee J_2) \text{ and} \\ (H_1 \times J_1) \wedge (H_2 \times J_2) &= (H_1 \wedge H_2) \times (J_1 \wedge J_2).\end{aligned}$$

The plenary subset $H \times J$ is closed under multiplication if and only if both H and J are closed. The closed plenary subsets of $A \times B$ form a lattice isomorphic with $\mathfrak{M} \times \mathfrak{B}$.

We next consider the questions, what sublattices does the lattice of the hypergroup have, and when is the lattice of $\{H\}$ among them? Theorem 9.1 contributes to the answer of the first part of the question, and the next two theorems to the second part.

¹¹ Dresher and Ore, pp. 714, 715.

¹² Dresher and Ore, p. 715, Theorem 2.

¹³ Garrett Birkhoff, "On the combinations of subalgebras," *Proceedings of the Cambridge Philosophical Society*, vol. 29 (1933), pp. 441-464, Theorem 18.1.

THEOREM 9.1. *Let H be a regular inversive hypergroup. The set of proper subhypergroups of H forms a lattice.*

Proof. If J and K are proper subhypergroups of H then $J \cup K$ is closed under multiplication, and contains the identity. If b is in $J \cup K$ then b^{-1} is in $J \cup K$, by the Lemma 1 following. Therefore $J \cup K$ is a proper subhypergroup.

The common part of J and K , $J \cap K$, is closed under multiplication, and contains the identity and the inverse of each of its elements. Therefore $J \cap K$ is a proper subhypergroup.

The operations of union and intersection are commutative and associative. Since the intersection of two proper subhypergroups is the largest contained in both, and the union the smallest containing both, $J \cup (K \cap J) = J$ and $(J \cup K) \cap J = J$. The following lemma then completes the proof.

LEMMA 1. *If H is an inversive regular hypergroup, then $c_e a_1 a_2 \cdots a_\eta$ implies that $c^{-1} e a_\eta^{-1} a_{\eta-1}^{-1} \cdots a_1^{-1}$.*

Proof by induction. Assume the conclusion valid for $\eta = 2$ and $\eta = \mu - 1$. It then follows for $\eta = \mu$. If $c_e a_1 a_2 \cdots a_{\mu-1} a_\mu$ then there is an element b in $a_1 a_2 \cdots a_{\mu-1}$ such that $c_e b a_\mu$, whence $c^{-1} e a_\mu^{-1} b^{-1}$. Since $b^{-1} e a_{\mu-1}^{-1} a_{\mu-2}^{-1} \cdots a_1^{-1}$, we have $c^{-1} e a_\mu^{-1} a_{\mu-1}^{-1} \cdots a_1^{-1}$. Thus Theorem 9.1 is proved.

THEOREM 9.2. *Let H be a hypergroup with a conjugation among its elements such that $\{H\}$ is regular and inversive. The proper appropriate subhypergroups of H form a lattice isomorphic with that of the proper subhypergroups of $\{H\}$.*

Proof. If J and K are proper and appropriate in H then $J \cap K$ is a proper appropriate subhypergroup, since it is closed under multiplication and the conjugation and contains all the identities and all the inverses of all its elements, as seen in Lemma 2. All the identities are in $J \cup K$. By Theorem 9.1 $\{J\} \cup \{K\}$ is a proper subhypergroup of $\{H\}$, whence by Lemma 2 there is a proper appropriate subhypergroup I of H such that $\{I\} = \{J\} \cup \{K\}$. By Lemma 3, since $\{J\} \cup \{K\}$ contains $\{J\}$ and $\{K\}$, I contains $J \cup K$. If i is an element in I then $\{i\}$ is in a product of the type $\{x_1\}\{x_2\} \cdots \{x_\eta\}$, where x_v is in J or K . This is only possible if there are elements x_v' , in J if $x_v \in J$, in K if $x_v \in K$, such that $i_e x_1' x_2' \cdots x_\eta'$. Thus every element $i_e I$ is in $J \cup K$, that is, $J \cup K = I$, a proper appropriate subhypergroup. As before, $J \cup (K \cap J) = J$ and $(J \cup K) \cap J = J$, and the lattice postulates are satisfied. The isomorphism follows from the formulas; $\{J \cup K\} = \{J\} \cup \{K\}$, $\{J \cap K\} = \{J\} \cap \{K\}$, which in turn follow from Lemmas 3 and 4.

LEMMA 2. Let H be a hypergroup with a conjugation among its elements such that $\{H\}$ is regular. For every proper subhypergroup K of $\{H\}$ there exists a proper appropriate subhypergroup J of H such that $\{J\} = K$. J contains all identities of H , and all the inverses of all its elements.

Proof. Let J be the set of all elements j which map upon the elements $\{j\}$ of K . J is appropriate and closed under multiplication. If e is an identity in H then $\{e\}$ is the identity of $\{H\}$, and therefore e is in J . If h^{-1} is an inverse of h with respect to e then $\{h^{-1}\}$ is the inverse of $\{h\}$ with respect to $\{e\}$. Therefore J contains with h all of its inverses. Thus J is proper and appropriate, and $\{J\} = K$.

LEMMA 3. If H is a hypergroup with a conjugation among its elements, and if J and K are appropriate subsets of H , then $J \supset K$ if and only if $\{J\} \supset \{K\}$.

Proof. If $J \supset K$ and $\{k\} \in \{K\}$, then $k \in K$, and so in J , and therefore $\{k\} \in \{J\}$. Therefore $\{J\} \supset \{K\}$. If $\{J\} \supset \{K\}$ and $k \in K$, then $\{k\} \in \{K\}$, whence $\{K\} \in \{J\}$, and so $k \in J$. Therefore $J \supset K$.

LEMMA 4. The union of two appropriate subsets is appropriate.

Proof by induction. Let h be an element of the union $J \vee K$ of two appropriate subsets. Then h is contained in a product $x_1 x_2 \cdots x_\mu$, where x_η is in either J or K . Let $h' \sim h$. If $\mu = 2$ then there exist $x_1' \sim x_1$, $x_2' \sim x_2$ such that $h' \in x_1' x_2'$. Suppose that for $\mu = \nu - 1$ there exist $x_\eta' \sim x_\eta$, $\eta = 1, 2, \dots, \nu - 1$, such that $h' \in x_1' x_2' \cdots x_{\nu-1}'$. Then a similar statement holds for $\mu = \nu$. For $h \in x_1 x_2 \cdots x_\nu$ implies that there is an element $b \in x_1 x_2 \cdots x_{\nu-1}$ such that $h \in b x_\nu$. If $h' \sim h$ then there exist $b' \sim b$, $x_\nu' \sim x_\nu$ such that $h' \in b' x_\nu'$. By hypothesis there exist $x_\eta' \sim x_\eta$, $\eta = 1, 2, \dots, \nu - 1$, such that $b' \in x_1' x_2' \cdots x_{\nu-1}'$. Therefore $h' \in b' x_\nu' \subset x_1' \cdots x_{\nu-1}' x_\nu'$. Since J and K are appropriate we have $x_\eta \in J$ implies $x_\eta' \in J$, and $x_\eta \in K$ implies $x_\eta' \in K$. Therefore $J \vee K$ is appropriate. The proof of Theorem 9.2 is now complete.

Let G be a group with a conjugation among its elements. If H is a subset of $\{G\}$ which is closed under multiplication, and if the set D in G which maps upon H is finite, then H is a subhypergroup of $\{G\}$. For D is closed under multiplication, and being finite, is an appropriate subgroup of G . Therefore $H = \{D\}$ is a subhypergroup of $\{G\}$.

If G is finite then the subhypergroups of $\{G\}$ form a lattice \mathcal{G} . For the closed subsets of $\{G\}$ form a lattice, and by the paragraph above each closed subset is a subhypergroup. Since, if D and F are appropriate subgroups of G ,

$\{D \vee F\} = \{D\} \vee \{F\}$ and $\{D \wedge F\} = \{D\} \wedge \{F\}$, the appropriate subgroups of G form a lattice isomorphic with \mathfrak{G} .

Let H be a regular hypergroup with a conjugation among its elements such that for every element b we have $b^{-1} \sim b$. Let J be a subset of $\{H\}$ which is closed under multiplication, and K be the set of elements of H which map upon the elements of J . Therefore K is closed under multiplication and the conjugation. With b it contains b^{-1} , and so it contains e . Therefore K is an appropriate proper subhypergroup of H .

If J and L are subhypergroups of H then $\{J \vee L\} = \{J\} \vee \{L\}$ and $\{J \wedge L\} = \{J\} \wedge \{L\}$. Thus we have

THEOREM 9.3. *Let H be a regular hypergroup with a conjugation among its elements such that for every b , $b \sim b^{-1}$. Then the subhypergroups of $\{H\}$ form a lattice isomorphic with that of the appropriate subhypergroups of H .*

We conclude with a condition that a group G be simple. Let S be a subgroup of the automorphisms of G . Let $b \sim c$ in G when $b \rightarrow c$ under an automorphism of S . Now $\{G\}_S$ is a finite regular inversive hypergroup, and the set of appropriate subgroups of G is a lattice isomorphic with that of the proper subhypergroups of $\{G\}_S$. A subgroup F is appropriate if and only if it is characteristic under S . If S is the group of inner automorphisms of G then the appropriate subgroups of G are the normal subgroups, and the lattice of the proper subhypergroups of $\{G\}_S$ is a B -lattice.¹⁴ We thus have the following theorem:

THEOREM 9.4. *If S is the group of inner automorphisms of the group G , then G is simple if and only if the partition hypergroup $\{G\}_S$ has no proper subhypergroups except E , the identity group.*

UNIVERSITY OF MINNESOTA.

¹⁴ Birkhoff, *loc. cit.*, Section II.

ON THE ALMOST PERIODIC BEHAVIOR OF MULTIPLICATIVE NUMBER-THEORETICAL FUNCTIONS.*

By E. R. VAN KAMPEN and AUREL WINTNER.

The purpose of the present paper is to develop criteria for the almost periodic behavior (B^λ) of multiplicative number-theoretical functions. In the particular case of what have been called strongly multiplicative functions, such criteria were recently¹ found for $\lambda = 1$ and $\lambda = 2$. However, the only representative of the classical number theoretical-functions in the class of strongly multiplicative functions is $\phi(n)/n$, where ϕ is Euler's function. Thus, there arises the question as to the possibility of a corresponding theory in the general case.

It will turn out that such a theory can be developed, although the situation then is essentially more involved. In fact, already the question of the almost periodicity (B^λ) of the factor functions, which belong to each of the prime numbers, must be discussed. Correspondingly, the preservation of almost periodicity (B^λ) on multiplication of a finite number of such factor functions requires especial care. The limit process which leads to the given multiplicative function is formally more involved than, though in principle not different from, the corresponding step in the strongly multiplicative case.

The results to be obtained may be illustrated by the sum, $\sigma(n)$, of the divisors of n . The result in this case will be that $\sigma(n)/n$ is almost periodic (B^λ) for arbitrarily large λ and has the Fourier expansion

$$\frac{\sigma(n)}{n} \sim \frac{6}{\pi^2} \sum_{m=1}^{\infty} \frac{c_m(n)}{m^2},$$

where the c 's denote the Ramanujan sums. But Ramanujan² has proved that

$$\sigma(n) = \frac{6}{\pi^2} n \sum_{m=1}^{\infty} \frac{c_m(n)}{m^2};$$

so that, if one divides by n , Ramanujan's *trigonometric series* turns out to be the *Fourier series* of the function to which it converges.

That Ramanujan's results do not imply any almost periodic behavior may

* Received November 15, 1939.

¹ M. Kac, E. R. van Kampen and Aurel Wintner, "Ramanujan sums and almost periodic functions," *American Journal of Mathematics*, vol. 62 (1940), pp. 107-114.

² S. Ramanujan, *Collected Papers* (Cambridge, 1927), pp. 179-199.

be illustrated by the following example: Ramanujan proves that if $d(n)$ denotes the number of divisors of n , then

$$d(n) = - \sum_{m=1}^{\infty} \frac{\log m}{m} c_m(n).$$

But this convergent trigonometrical series cannot be the Fourier series (B) of the function, $d(n)$, which it represents. In fact,

$$\frac{1}{n} \sum_{m=1}^n d(m) \sim \log n, \quad n \rightarrow \infty, \quad (\text{Dirichlet})$$

implies that the mean value of $d(n)$ is $+\infty$; so that $d(n)$ cannot be almost periodic (B).

Incidentally, the results to be obtained are, in contrast to the results of Ramanujan, independent of the prime number theorem.

It should be mentioned that, while Theorems IV and VI below may, with straightforward modification of proof and wording, be transferred to the case where multiplicative functions are replaced by additive functions, essential complications seem to arise in connection with the corresponding analogue to Theorem V below (if $\lambda \neq 1$).

By a function $f(n)$ will be meant a sequence in which n runs through all positive integers. The average $M\{f\} = M\{f(n)\}$ of an f is defined as the limit ($n \rightarrow \infty$) of the arithmetical mean of the n numbers $f(1), \dots, f(n)$, if this limit exists. And $\bar{M}\{f\} = \bar{M}\{f(n)\}$ will denote the upper limit ($\leq +\infty$) of this arithmetical mean, if $f \geq 0$.

By a multiplicative function $f(n)$ is meant a sequence $f(1), f(2), \dots$ of numbers for which

$$f(n_1 n_2) = f(n_1) f(n_2) \text{ whenever } (n_1, n_2) = 1; \text{ hence, } f(1) = 1$$

unless $f(n) = 0$ for every n (this trivial case will be excluded).

If there exists a fixed prime number p^* such that $f(p^k) = 1$ for every k and for every prime number p distinct from p^* , the function $f(n)$ will be called a prime multiplicative function (belonging to the prime number p^*). It is clear that if p_m denotes the m -th prime number and $f_m(n)$ an arbitrary prime multiplicative function belonging to p_m , then

$$(1) \quad f(n) = \prod_{m=1}^{\infty} f_m(n)$$

defines a multiplicative function $f(n)$, all but a finite number of the factors of the infinite product being 1 for a fixed n . Conversely, every given multi-

plicative function f determines a unique sequence $\{f_m\}$ of prime multiplicative functions f_m by means of which f is representable in the form (1). In fact,

$$(2) \quad f_m(n) = f(p_m^k) \text{ if } p_m^k | n \text{ but } p_m^{k+1} \nmid n; \quad (f(1) = 1).$$

In what follows, $g(n)$ will denote an arbitrary function which satisfies, for a fixed prime p , the requirement that

$$(3) \quad g(n) = g(p^k) \text{ if } p^k | n \text{ but } p^{k+1} \nmid n.$$

Condition (3) is satisfied by every prime multiplicative function belonging to p , but not only by these functions; in fact, (3) is possible also when $g(n)$ vanishes for $n = 1$, without vanishing for every n .

THEOREM I. *The average $M\{g\}$ of a function (3) exists if and only if the series*

$$(4) \quad \sum_{k=0}^{\infty} p^{-k} g(p^k) \text{ is convergent,}$$

in which case

$$(5) \quad M\{g\} = (1 - p^{-1}) \sum_{k=0}^{\infty} p^{-k} g(p^k).$$

Remark. It will be clear from the proof that (5) holds, if $g \geq 0$, also when the series (4) is divergent (in which case $M\{g\} = +\infty$).

Proof. Let a_i denote the arithmetical mean of the p^i numbers $g(1), \dots, g(p^i)$, where i is an arbitrary non-negative integer and p the prime number belonging to g . Then, by (3),

$$(6) \quad a_i = (1 - p^{-1}) \sum_{k=0}^i p^{-k} g(p^k) + p^{-i-1} g(p^i).$$

Hence, for every $i > 0$,

$$a_i - a_{i-1} = p^{-i} (g(p^i) - g(p^{i-1})); \quad a_0 = g(1),$$

and so

$$(7) \quad p^{-i} g(p^i) = p^{-i} a_0 + \sum_{k=1}^i p^{-i+k} (a_k - a_{k-1}); \quad (i = 0, 1, \dots).$$

Suppose first that $M\{g\}$ exists. Then, in particular,

$$a_i \rightarrow M\{g\}, \text{ and so } a_i - a_{i-1} \rightarrow 0, \text{ as } i \rightarrow \infty.$$

Consequently, application to (7) of a standard lemma concerning linear summation methods shows, that $p^{-i} g(p^i) \rightarrow 0$ as $i \rightarrow \infty$. Hence, (4) and (5) follow from (6).

In order to prove the sufficiency of (4) for the existence of $M\{g\}$, let

$$n = \sum_{i=0}^m q_i p^i, \text{ where } 0 \leq q_i < p; \quad m = m(n), q_i = q_i(n),$$

be the p -adic representation of an arbitrary $n > 0$. Then, by (3) and (6),

$$\frac{1}{n} \sum_{i=1}^n g(i) = \frac{\sum_{i=0}^m q_i p^i a_i}{\sum_{i=1}^m q_i p^i}.$$

It follows, therefore, from the standard lemma on linear summation methods, used before, that in order to prove the existence of $M\{g\}$, it is sufficient to assure that the sequence a_1, a_2, a_3, \dots (which merely is a subsequence of the sequence defining $M\{g\}$) has a limit $\neq \pm \infty$. But (6) and the assumption (4) clearly imply the existence of $\lim a_i (\neq \pm \infty)$; so that the proof is complete.

THEOREM II. *A function g which satisfies (3) (and so, in particular, a prime multiplicative function belonging to a prime p) is almost periodic (B^λ) for a given positive $\lambda (\geq 1)$ if and only if ³*

$$(8) \quad \sum_{k=0}^{\infty} p^{-k} |g(p^k)|^\lambda < \infty.$$

(This implies, for $\lambda = 1$, the curious fact that $g(n)$ is almost periodic (B) whenever so is $|g(n)|$).

It is understood that if $\lambda < 1$ in Theorem II (so that there is no Hölder-Minkowski inequality and, correspondingly, no natural metric in the B^λ -space), then $M\{g\}$ need not exist, and so, in particular, $g(n)$ need not have a Fourier expansion.

Proof. If $g(n)$ is almost periodic (B^λ), so is $|g(n)|$; so that $M\{|g|^\lambda\}$ exists. Consequently, application of Theorem I to the function $|g(n)|^\lambda$ shows, that (8) is a necessary condition for the almost periodicity (B^λ) of g .

In order to prove the converse, define, in terms of any given function g , for every positive integer j a function g^j , by placing

$$(9) \quad g^j(n) = g(n) \text{ if } 1 \leq n \leq p^j, \quad g^j(n + p^j) = g^j(n) \text{ for every } n.$$

Then it is clear that (3) remains valid if one replaces $g(n)$ by the non-negative function $|g(n) - g^j(n)|^\lambda$ of n for a fixed j . Hence, the Remark which follows Theorem I implies that

$$M\{|g - g^j|^\lambda\} = (1 - p^{-1}) \sum_{k=0}^{\infty} p^{-k} |g(p^k) - g^j(p^k)|^\lambda.$$

³ This result is closely related to a construction due to O. Toeplitz, "Ein Beispiel zur Theorie der fastperiodischen Funktionen," *Mathematische Annalen*, vol. 98 (1927), pp. 281-295.

On letting here $j \rightarrow \infty$, one sees from (9) that, if (8) is satisfied,

$$M\{|g - g^j|^\lambda\} \rightarrow 0 \text{ as } j \rightarrow \infty.$$

Since every g^j is, by (9), a periodic function of n , it follows that g is almost periodic (B^λ). This completes the proof of Theorem II.

THEOREM III. *A function $g(n)$ which satisfies (3) is almost periodic (B) if and only if*

$$(10) \quad \sum_{k=0}^{\infty} p^{-k} |g(p^k)| < \infty,$$

in which case the Fourier expansion of $g(n)$ is

$$(11) \quad g(n) \sim M\{g\} + \sum_{j=1}^{\infty} c_{p^j}(n) \sum_{k=j}^{\infty} \frac{g(p^k) - g(p^{k-1})}{p^k},$$

where the constant term is

$$(12) \quad M\{g\} = (1 - p^{-1}) \sum_{k=0}^{\infty} p^{-k} g(p^k), \text{ by (5),}$$

and the $c_{p^j}(n)$ denote the Ramanujan sums belonging to those indices m which are powers of p :

$$(13) \quad c_m(n) = \sum_k \exp\left(2\pi i \frac{k}{m} n\right) \\ = \sum_k \cos 2\pi \frac{k}{m} n, \text{ where } 1 \leq k \leq m \text{ and } (k, m) = 1.$$

In particular, $g(n)$ is limit-periodic (grenzperiodisch), since the Fourier constants belonging to (circular) irrational frequencies all vanish.

Remark. Assuming that (8) is satisfied for $\lambda = 2$, one sees from (12) and (13) that the Parseval relation belonging to (11) is

$$(14) \quad \sum_{k=0}^{\infty} \left(1 - \frac{1}{p}\right) \frac{|g(p^k)|^2}{p^k} \\ = \left| \left(1 - \frac{1}{p}\right) \sum_{k=0}^{\infty} \frac{g(p^k)}{p^k} \right|^2 + \sum_{j=1}^{\infty} (p^j - p^{j-1}) \left| \sum_{k=j}^{\infty} \frac{g(p^k) - g(p^{k-1})}{p^k} \right|^2,$$

an identity which can, of course, be verified directly.

Proof. Since (10) is the particular case $\lambda = 1$ of the criterion (8) of Theorem II, only the explicit form (11) of the Fourier expansion needs a proof. To this end, one can readily verify from the definitions (12), (13) and (9), that (11) is certainly true if g is replaced by the periodic function g^j (where j is arbitrarily fixed); in fact, (11) follows for $g = g^j$ by straightforward trigonometrical interpolation. Since, by the proof of Theorem II, one has $M\{|g - g^j|\} \rightarrow 0$ as $j \rightarrow \infty$, it follows that (11) holds for any g .

Since the preceding results concern an arbitrary function which satisfies (3), they are applicable to every prime multiplicative function, and so to any of the factors (2) of an arbitrary multiplicative function (1). In what follows, there will be established natural analogues to Theorems I-III for the case where prime multiplicative functions $g(n) = f_m(n)$ are replaced by arbitrary multiplicative functions $f(n)$. It seems to be hard to replace Theorems IV, V, VI below by theorems which are of the same sharpness as the corresponding Theorems I, II, III above.

It will be convenient to associate with every multiplicative function $f(n)$ another multiplicative function $f_*(n)$, which is defined by

$$(15) \quad f_*(p^k) = \frac{f(p^k) - f(p^{k-1})}{p^k}.$$

Then, since also $f(n)$ is multiplicative,

$$(16) \quad f(n) = \sum_{d|n} df_*(d).$$

THEOREM IV. *The average $M\{f\}$ of a multiplicative function $f(n)$ exists whenever*

$$(17) \quad \sum_{n=1}^{\infty} |f_*(n)| < \infty,$$

in which case

$$(18) \quad M\{f\} = \sum_{n=1}^{\infty} f_*(n).$$

Remark. It will be clear from the proof that (18) holds, if $f_*(n) \geq 0$, also when the series (17) is divergent (in which case $M\{f\} = +\infty$).

Proof. It is seen from (16) that, for every $n \geq 1$,

$$\sum_{k=1}^n f(k) = \sum_{k=1}^n \left[\frac{n}{k} \right] kf_*(k).$$

Hence,

$$\sum_{k=1}^n f(k) = n \sum_{k=1}^n f_*(k) + O\left(\sum_{k=1}^n k |f_*(k)|\right)$$

as $n \rightarrow \infty$. Since (17) implies that $\frac{1}{n} \sum_{k=1}^n k |f_*(k)| \rightarrow 0$, it follows that

$$\sum_{k=1}^n f(k) = n \sum_{k=1}^n f_*(k) + o(n).$$

This completes the proof of Theorem IV.

It will be convenient to extend the class of an arbitrary multiplicative function f in the same way as condition (3) extends the class of prime

multiplicative functions. The extension in question may be defined by the requirement that the given $f(n)$ admits of a factorization (1) in which a factor f_m need not be multiplicative but merely such that condition (3) is satisfied by $g = f_m$ and $p = p_m$, where $m = 1, 2, \dots$. Since $f_m(1)$ need not be 1, it must, of course, be assumed that the product $\prod_{m=1}^{\infty} f_m(1)$ is convergent and remains convergent if one omits a finite number of its factors.

If these conditions are satisfied, f will be called a generalized multiplicative function. For the proof of Theorem V below, those and only those generalized multiplicative functions will be needed for which $f_m(1) = 1$ holds for every m with the exception of one value, say $m = r$, for which $f_r(1) = 0$.

For a generalized multiplicative function f , let f_* denote the generalized multiplicative function

$$(19) \quad f_*(n) = \prod_{m=1}^{\infty} f_{m*}(n), \text{ where } f_{m*}(1) = f_m(1), f_{m*}(p^k) = \frac{f_m(p^k) - f_m(p^{k-1})}{p^k},$$

and every f_{m*} is prime multiplicative. It is clear that this definition of f_* reduces to the definition (15) if the generalized multiplicative function f is multiplicative.

THEOREM IV bis. *Theorem IV holds for generalized multiplicative functions f also.*

This is readily seen from the proof of Theorem IV and from the definitions of the generalized multiplicative functions f, f_* .

The following considerations will be based on an auxiliary lemma.

LEMMA. *The product*

$$(20) \quad f(n) = \prod_{r=1}^m f_r(n)$$

of a finite number of prime multiplicative functions f_1, \dots, f_m which belong to distinct prime numbers p_1, \dots, p_m is almost periodic (B^λ) for a fixed $\lambda \geq 1$ whenever each of the functions f_1, \dots, f_m is almost periodic (B^λ) for this λ .

Proof. For every $r (= 1, \dots, m)$, define two prime multiplicative functions u_r, v_r of n by placing

$$v_r(n) = \text{Max} (1, |f_r(n)|), \quad u_r(n) = f_r(n)/v_r(n);$$

so that

$$|u_r(n)| \leq 1 \leq v_r(n)$$

and

$$f_r(n) = u_r(n)v_r(n).$$

Since u_r is a bounded function of n , criterion (8) of Theorem II is satisfied by $g = u_r$ for every exponent, and so u_r is almost periodic (B^κ) for every κ . On the other hand, since f_r is supposed to be almost periodic (B^λ) for a fixed λ , Theorem II assures that (8) is satisfied by $g = f_r$, and so, in view of the definition of v_r , by $g = v_r$ also; so that v_r is almost periodic (B^λ) by Theorem II. Since the product f of the m functions f_r is the product of the $m + m$ functions u_r, v_r , where the u_r are almost periodic (B^κ) for arbitrarily large κ , and since $\lambda \geq 1$ by assumption, it follows that the proof of the Lemma will be complete if one shows that the product

$$(21) \quad v(n) = \prod_{r=1}^m v_r(n)$$

of the m almost periodic (B^λ) functions v_r is almost periodic (B^λ). And this may be shown by induction from m to $m + 1$, as follows:

Suppose that v is already known to be almost periodic (B^λ). Let w be a prime multiplicative function which belongs to a prime number p distinct from the primes p_1, \dots, p_m to which the factors v_1, \dots, v_m of v belong. Suppose finally that w is (as are these factors v_r of v) not less than 1 and almost periodic (B^λ). The induction from m to $m + 1$ requires to prove that the product vw is almost periodic (B^λ). To this end, let $w^j = w^j(n)$ denote the function which one obtains by applying the definition (9) to $g = w$, where $j = 1, 2, \dots$. Then w^j is periodic, hence almost periodic (B^κ) for every κ , and so the product vw^j is almost periodic (B^λ), since $\lambda \geq 1$. Hence, the proof of the almost periodicity (B^λ) of vw will be complete if one shows that

$$M\{d^j\} \rightarrow 0 \text{ as } j \rightarrow \infty,$$

where $d^j = d^j(n)$ denotes the function

$$d^j = |vw - vw^j|; \quad (j = 1, 2, \dots).$$

But it is clear from the definitions of v, w and w^j , that d^j is for every fixed j a generalized multiplicative function in the sense defined before Theorem IV bis. Hence, by Theorem IV bis, it is sufficient to show [cf. (17)-(18)] that

$$(22) \quad \sum_{n=1}^{\infty} |d^{j*}(n)|$$

is convergent for a fixed j and that

$$|M\{d_j\}| = \left| \sum_{n=1}^{\infty} d^{j*}(n) \right| \leq \sum_{n=1}^{\infty} |d^{j*}(n)| \rightarrow 0 \text{ as } j \rightarrow \infty,$$

where $d^{j*}(n)$ is obtained by applying the definition (19) to $f = d^j$. And this may be proved as follows:

Since v_1, \dots, v_m and w are prime multiplicative and almost periodic (B^λ), Theorem II assures that condition (8) is satisfied by any of the $m+1$ functions $g = w, v_1, \dots, v_m$; so that, by the definition (19) of the asterisk symbol.

$$(23) \quad \sum_{n=1}^{\infty} |d^0(n)| < \infty,$$

if $d^0 = d^0(n)$ denotes the multiplicative function

$$d^0 = (vw)^\lambda = (v_1 \cdots v_m w)^\lambda; \quad \text{cf. (21).}$$

But it is clear from (19) and from the definitions of w^j and d^j , where $j > 0$, that the series (23) is a majorant of (22), also if one replaces by zeros those terms of (23) which are not divisible by the $(j+1)$ -th power of p . Hence, it is clear from (23) that the series (22) is convergent for every j and has a value which tends to 0 as $j \rightarrow \infty$.

This completes the proof of the Lemma.

Remark. It may be mentioned that also the converse of the Lemma is true, i. e., that for the almost periodicity (B^λ) of a finite product of prime multiplicative functions belonging to different primes it is not only sufficient but also necessary that each of these prime multiplicative functions be almost periodic (B^λ), where $\lambda \geq 1$ is arbitrarily fixed. This converse of the Lemma is not needed in what follows, so that the proof will be omitted.

THEOREM V. *If $\lambda \geq 1$ is fixed and f is a real non-negative multiplicative function, then f is almost periodic (B^λ) whenever*

$$(24) \quad \sum_{n=1}^{\infty} |f^\lambda(n)| < \infty.$$

It is understood that by f^λ is meant the function which belongs to f^λ in the same way as the multiplicative function f_* defined by (15) belongs to f , and that $f^\lambda = f^\lambda(n)$ denotes the λ -th power of $f = f(n)$.

Remark. In view of the Remark which follows Theorem IV, condition (24) is necessary as well for the almost periodicity (B^λ) of f in case $f_* \geq 0$.

Proof. It is clear from the definition (19) that if (24) is satisfied by the non-negative multiplicative function f , then it is also satisfied by each of its prime multiplicative factors $f_m \geq 0$. Thus, condition (8) is satisfied by $g = f_1, f_2, \dots$, and so Theorem II assures the almost periodicity (B^λ) of any of these prime multiplicative functions. It follows therefore from the preceding Lemma that the multiplicative function

$$(25) \quad h_m(n) = \prod_{r=1}^m f_r(n)$$

of n is almost periodic (B^λ) for every m . Hence, in order to complete the proof of Theorem V, it would be sufficient to prove that

$$(26) \quad M\{|f - h_m|^\lambda\} \rightarrow 0 \text{ as } m \rightarrow \infty.$$

However, it will be convenient to carry out the limit process leading from (25) to (1) in another manner, namely by means of a pair of auxiliary functions $u = u(n)$, $v = v(n)$ which are defined in terms of the given function $f = f(n)$ as follows:

Both functions $u(n)$, $v(n)$ are multiplicative, and

$$(27) \quad u(p^k) = \text{Min}(1, f(p^k)); \quad v(p^k) = \text{Max}(1, f(p^k))$$

for every prime p and for every $k \geq 1$. Since $f \geq 0$ by assumption, it is clear from (27) that

$$(28) \quad 0 \leq u \leq 1 \leq v;$$

while

$$(29) \quad f = uv$$

for every n . Furthermore, from (27) and (15),

$$(30) \quad |u \cdot| \leq |f \cdot|; \quad |v \cdot| \leq |f \cdot|.$$

By u_m and v_m will be meant the prime multiplicative functions which belong to the m -th prime number $p = p_m$ in the same way as f_m in (1) and (2) belongs to f ; so that

$$(31) \quad u(n) = \prod_{m=1}^{\infty} u_m(n); \quad v(n) = \prod_{m=1}^{\infty} v_m(n).$$

It is clear from (30) that the assumption (24) remains satisfied if one replaces f by u or by v . Since it was shown before that each of the partial products (25) of the infinite product (1) is almost periodic (B^λ) if f satisfies (24), it follows that each of the partial products of either of the infinite products (31) is almost periodic (B^λ). Hence, in order to prove that u and v are almost periodic (B^λ), it is sufficient to show that

$$(32) \quad \bar{M}\{|u - x_m|^\lambda\} \rightarrow 0; \quad \bar{M}\{|v - y_m|^\lambda\} \rightarrow 0 \text{ as } m \rightarrow \infty,$$

where $x_m = x_m(n)$, $y_m = y_m(n)$ denote the multiplicative functions

$$(33) \quad x_m = \prod_{r=1}^m u_r; \quad y_m = \prod_{r=1}^m v_r.$$

But if both functions u, v are known to be almost periodic (B^λ), then also

their product is almost periodic (B^λ), since u is bounded, by (28). It follows therefore from (29) that the proof of Theorem V will be complete if one verifies (32).

In order to prove (32), notice that, by (27), (28) and (31),

$$(34) \quad 1 \geq x_1 \geq x_2 \geq x_3 \geq \dots \geq u \geq 0; \quad 1 \leq y_1 \leq y_2 \leq y_3 \leq \dots \leq v$$

for every n . Furthermore, (30) and (24) assure that (17) is satisfied by any of the functions $f = u^\lambda, v^\lambda, x_m^\lambda, y_m^\lambda$; so that, by Theorem IV, the average of any of these functions exists. And these averages satisfy, in view of (18), the limit relations

$$(35) \quad M\{x_m^\lambda\} \rightarrow M\{u^\lambda\}; \quad M\{y_m^\lambda\} \rightarrow M\{v^\lambda\} \text{ as } m \rightarrow \infty.$$

But it is clear from (34) and (35) that the proof of (32), hence also the proof of Theorem V, will be complete if one verifies the following elementary lemma (which has nothing to do with multiplicative functions):

Lemma. If there exists a finite average $M\{f\}$ for the λ -th power ($\lambda \geq 1$) of each of the real non-negative functions $f(n) = F(n); F_1(n), F_2(n), \dots$ and if, for every fixed m ,

(36) either $F_m(n) \leq F(n)$ for every n or $F_m(n) \geq F(n)$ for every n , then the limit relation

$$(37) \quad M\{F_m^\lambda\} \rightarrow M\{f^\lambda\}, \quad m \rightarrow \infty,$$

implies that

$$(38) \quad \bar{M}\{|F - F_m|^\lambda\} \rightarrow 0, \quad m \rightarrow \infty.$$

In order to prove this Lemma, notice that, since $\lambda \geq 1$,

$$(1 - t)^\lambda \leq 1 - t^\lambda \text{ for } 0 \leq t \leq 1.$$

Hence, it is clear from (36) that

$$|F - F_m|^\lambda \leq |F^\lambda - F_m^\lambda|.$$

Consequently, (38) follows from (37) in view of (36).

This completes the proof of Theorem V.

Theorem V may be refined by exhibiting a sequence of almost periodic functions (B^λ) which are explicitly defined in terms of f and tend to f with reference to the metric of the space (B^λ):

THEOREM V bis. *If a real non-negative multiplicative function f satisfies (24) for a fixed $\lambda \geq 1$, then the products (1), (25) are almost periodic (B^λ) and satisfy (26).*

Proof. It was shown in the Proof of Theorem V that each of the real

non-negative functions $h_m; x_m, y_m; u, v; f$ of n is almost periodic (B^λ). Furthermore, (25), (27) and (33) imply that $h_m = x_m y_m$; so that, by (29),

$$f - h_m = (u - x_m)v + (v - y_m)x_m.$$

Hence, if $\lambda > 1$, it is seen from Minkowski's inequality that in order to prove (26), it is sufficient to show that

$$M\{|(u - x_m)v|^\lambda\} \rightarrow 0; \quad M\{|(v - y_m)x_m|^\lambda\} \rightarrow 0, \quad (m \rightarrow \infty),$$

where $u; x_1, x_2, \dots$ are, by (28) and (34), uniformly bounded functions of n . Consequently, if $\lambda > 1$, the assertion (26) is, in view of Hölder's inequality and of the almost periodicity (B^λ) of v , implied by

$$M\{|u - x_m|\} \rightarrow 0; \quad M\{|v - y_m|^\lambda\} \rightarrow 0, \quad (m \rightarrow \infty),$$

a pair of relations which obviously imply (26) in the limiting case $\lambda = 1$ also. Since this pair of relations is, in view of the uniform boundedness of the functions $u - x_1, u - x_2, \dots$ of n , equivalent to (32), the proof of Theorem V bis is complete.

THEOREM VI. *A multiplicative function $f \geq 0$ is almost periodic (B) whenever*

$$(39) \quad \sum_{m=1}^{\infty} |f_*(n)| < \infty,$$

where $f_*(n)$ is the multiplicative function defined by

$$f_*(p^k) = \frac{f(p^k) - f(p^{k-1})}{p^k}.$$

The Fourier expansion of f then is

$$(40) \quad f(n) \sim \sum_{m=1}^{\infty} a_m c_m(n), \quad \text{where } a_m = \sum_{l=1}^{\infty} f_*(ml),$$

$c_m(n)$ denoting the Ramanujan sum (13). In particular, f is limit-periodic (grenzperiodisch).

Proof. It is clear from the assumptions of Theorem VI that the assumptions of Theorem III are satisfied by each of the prime multiplicative functions $g = f_1, f_2, \dots$ which occur in the factorization (1) of f . It follows therefore from (11) and (12) that (40) is true if f is replaced by any of the functions f_1, f_2, \dots . Since (13) is known to be a multiplicative function of m for every fixed n , it follows that the series (40) belonging to the finite product $f = h_m = f_1 f_2 \dots f_m$ may be obtained by a formal multiplication of the m Fourier series (40) which belong to $f = f_1, f_2, \dots, f_m$, respectively. But the resulting formal product is identical with the Fourier series of the function

$h_m = f_1 f_2 \cdots f_m$, as seen by a repeated application (for $\lambda = 1$) of the Lemma which precedes Theorem Vbis. Consequently, (40) is true if the infinite product (1) is replaced by the finite product (25), where m is arbitrary. Hence, on applying (26) for $\lambda = 1$, one sees that (40) holds for the infinite product (1) also. This completes the proof of Theorem VI.

The above results will now be applied to the case of classical multiplicative functions, involving Euler's ϕ -function and the sum, $\sigma_\alpha(n)$, of the α -th powers of the divisors of n ; so that

$$(41) \quad \sigma_\alpha(n) = \sum_{d|n} d^\alpha, \text{ i. e., } \sigma_\alpha(p^k) = \frac{p^{(k+1)\alpha} - 1}{p^\alpha - 1}$$

(where it is understood that $\sigma_0(p^k)$ denotes the limit ($= k + 1$) of $\sigma_\alpha(p^k)$ as $\alpha \rightarrow +0$; so that the function $\sigma_0(n)$ represents the number of the divisors of n).

For sake of shortness, a function will be called almost periodic (B^∞) if it is almost periodic (B^λ) for arbitrarily large λ .

(i) *The multiplicative function $f(n) = \sigma_\alpha(n)/n^\alpha$ is almost periodic (B^∞) for every $\alpha > 0$ and has the Fourier expansion*

$$(42) \quad \frac{\sigma_\alpha(n)}{n^\alpha} \sim \zeta(\alpha + 1) \sum_{m=1}^{\infty} \frac{c_m(n)}{m^{\alpha+1}}, \quad (\alpha > 0);$$

while $M\{\sigma_0\} = +\infty$.

Proof. In order to prove the almost periodicity (B^∞), it is sufficient to show that the criterion (24) of Theorem V is satisfied for every positive λ . But if $f(n) = \sigma_\alpha(n)/n^\alpha$, then, by (41),

$$(43) \quad f^\lambda(p^k) = \frac{(p^{(k+1)\alpha} - 1)^\lambda}{(p^\alpha - 1)^\lambda p^{k\alpha\lambda}}; \text{ so that } f^\lambda(p^k) = \frac{(p^{(k+1)\alpha} - 1)^\lambda - p^{a\lambda}(p^{k\alpha} - 1)^\lambda}{(p^\alpha - 1)^\lambda p^{a\lambda k + k}},$$

by (15). Hence, it is easy to see that $f^\lambda > 0$; so that, on applying to the series (24) the Euler factorization, one sees that the criterion of Theorem V requires that

$$\prod_p \sum_{k=0}^{\infty} \frac{(p^{(k+1)\alpha} - 1)^\lambda - p^{a\lambda}(p^{k\alpha} - 1)^\lambda}{(p^\alpha - 1)^\lambda p^{a\lambda k + k}} < \infty \text{ for every } \lambda > 0.$$

But this product of sums may be rewritten as

$$\prod_p \sum_{k=0}^{\infty} \frac{(1 - p^{-a(k+1)})^\lambda - (1 - p^{-ak})^\lambda}{(1 - p^{-a})^\lambda p^k} = \prod_p \sum_{j=1}^{\infty} \frac{(p-1)(1 - p^{-aj})^\lambda}{(1 - p^{-a})^\lambda p^j},$$

and is therefore of the form

$$\begin{aligned} \prod_p \left(\frac{p-1}{p} + \frac{(p-1)(1 - p^{-a})^\lambda}{p^2} + O(p^{-2}) \right) \\ = \prod_q (1 + \lambda p^{-1-a} + O(p^{-1-2a}) + O(p^{-2})). \end{aligned}$$

And this product is absolutely convergent for every $\lambda > 0$ and for every $\alpha > 0$, since $\sum p^{-1-\epsilon} < \infty$ for every $\epsilon > 0$.

This completes the proof of the almost periodicity (B^∞) of (41) for every $\alpha > 0$. Since application of (43) for $\lambda = 1$ shows that $f_\bullet(p^k) = p^{-\alpha k-k}$, one has, for $m = 1, 2, \dots$,

$$\sum_{l=1}^{\infty} f_\bullet(ml) = \sum_{l=1}^{\infty} (ml)^{-1-\alpha} = \zeta(\alpha+1)/m^{\alpha+1};$$

so that (40) reduces to (42).

The calculations involved become shorter in case of Euler's ϕ -function, in which case the result is as follows:

(ii) Both functions $\phi(n)/n$, $n/\phi(n)$ are almost periodic (B^∞). Their Fourier expansions are

$$(44_1) \quad \frac{\phi(n)}{n} \sim \frac{1}{\zeta(2)} \sum_q c_q(n) \prod_{p|q} \frac{p^2}{p^2-1}; \quad (44_2) \quad \frac{n}{\phi(n)} \sim \frac{\zeta(2)}{\zeta(3)} \sum_q c_q(n) \prod_{p|q} \frac{1-p^{-2}}{1-p^{-3}},$$

where the summation index q runs through the quadratfrei positive integers.

Proof. Since $\phi(p^k) = p^k - p^{k-1}$, application of the definition (15) to the λ -th powers of the respective functions $f(n) = \phi(n)/n$ and $f(n) = n/\phi(n)$ gives

$$(45_1) \quad f_\bullet^\lambda(p) = \frac{(p-1)^\lambda - p^\lambda}{p^{\lambda+1}}; \quad f_\bullet^\lambda(p^k) = 0 \text{ if } k > 1,$$

$$(45_2) \quad f_\bullet^\lambda(p) = \frac{p^\lambda - (p-1)^\lambda}{(p-1)^\lambda p}; \quad f_\bullet^\lambda(p^k) = 0 \text{ if } k > 1.$$

Hence, the criterion (24) of Theorem V requires, for a fixed λ , that

$$\prod_p (1 + |A_p(\lambda)|) < \infty, \text{ i. e., that } \sum_p |A_p(\lambda)| < \infty,$$

where

$$A_p(\lambda) = \frac{(p-1)^\lambda - p^\lambda}{p^{\lambda+1}} \text{ and } A_p(\lambda) = \frac{p^\lambda - (p-1)^\lambda}{(p-1)^\lambda p},$$

respectively. Since it is clear from $\sum p^{-1-\epsilon} < \infty$ that $\sum |A_p(\lambda)| < \infty$ holds for every $\lambda > 0$ in both cases, the almost periodicity (B^∞) follows. Finally, application of (44₁) and (44₂) for $\lambda = 1$ gives

$$f_\bullet(p) = -\frac{1}{p^2}, f_\bullet(p^k) = 0, k > 1 \text{ and } f_\bullet(p) = \frac{1}{p(p-1)}, f_\bullet(p^k) = 0, k > 1,$$

respectively; so that (44₁) and (44₂) readily follow from (40).

ON UNIFORMLY ALMOST PERIODIC MULTIPLICATIVE AND ADDITIVE FUNCTIONS.*

By E. R. VAN KAMPEN.

In this note conditions are established for certain multiplicative or additive functions to be uniformly almost periodic¹ (u. a. p.).

A subscript p on the symbols Π or Σ will denote a product or sum over all primes, except that sometimes (explicitly) a finite number of primes will be excluded.

A multiplicative function f is an arithmetical function $f = f(n)$, $n = 1, 2, \dots$, which satisfies

$$(1) \quad f(n_1 n_2) = f(n_1) f(n_2) \text{ if } (n_1, n_2) = 1, \quad [f(1) = 1].$$

Such a function may be written in the form

$$(2) \quad f(n) = \Pi_p f_p(n), \quad (n = 1, 2, \dots),$$

where $f_p(n)$ is for a fixed prime p defined by

$$(3) \quad f_p(n) = f(p^k) \text{ if } p^k | n \text{ and } p^{k+1} \nmid n.$$

The product in (2) clearly is a finite product for every fixed n .

An additive function g is an arithmetical function $g = g(n)$, $n = 1, 2, \dots$, which satisfies

$$(4) \quad g(n_1 n_2) = g(n_1) + g(n_2) \text{ if } (n_1, n_2) = 1, \quad [g(1) = 0].$$

Such a function may be written in the form

$$(5) \quad g(n) = \Sigma_p g_p(n), \quad (n = 1, 2, \dots),$$

where $g_p(n)$ is, for a fixed prime p , defined by

$$(6) \quad g_p(n) = g(p^k) \text{ if } p^k | n \text{ and } p^{k+1} \nmid n.$$

Thus the sum in (5) is a finite sum for every fixed n .

The main result may be formulated as follows:

THEOREM 1. *An additive function $g(n)$ [real-valued multiplicative function $f(n)$] is u. a. p. if and only if the sum representation (5) [the product*

* Received November 30, 1939.

¹ Conditions for such functions to belong to a class (B^λ) of Besicovitch almost periodic functions are investigated in: E. R. van Kampen and Aurel Wintner, *American Journal of Mathematics*, vol. 62 (1940), pp. 613-626.

representation (2)] is uniformly convergent and each summand g_p [factor f_p] is u. a. p.

The sufficiency of the above conditions is, of course, evident from the elementary properties of u. a. p. functions, also for complex valued multiplicative functions. On expressing the conditions in analytical terms, one obtains the following theorems:

THEOREM 2. *An additive function g is u. a. p. if and only if the series*

$$(7) \quad \sum_p \text{l. u. b.}_k |g(p^k)| \text{ is convergent,}$$

and the limit

$$(8) \quad \alpha_p = \lim g(p^k) \text{ exists for every prime } p, \quad (k \rightarrow \infty).$$

It will be clear from the proof that if $g(n)$ is u. a. p., then the unique u. a. p. extension of $g(n)$ to non-positive values of n may be obtained by placing $g(-n) = g(n)$ and $g(0) = \sum_p \alpha_p$.

THEOREM 3. *A multiplicative function f is u. a. p. if, and in case of real-valued functions only if, the series*

$$(9) \quad \sum_p \text{l. u. b.}_k |f(p^k) - 1| \text{ is convergent,}$$

and the limit

$$(10) \quad \beta_p = \lim f(p^k) \text{ exists for every } p, \quad (k \rightarrow \infty).$$

And one sees easily that the unique u. a. p. extension of $f(n)$ for non-positive n may be obtained by placing $f(-n) = f(n)$ and $f(0) = \prod_p \beta_p$.

The proof of Theorem 2 is simpler than the proof of Theorem 3. One could reduce Theorem 2 to a special case of Theorem 3 by considering the multiplicative functions $\exp(\Re g(n))$ and $\exp(\Im g(n))$, where $g(n)$ is additive. The, apparently difficult, complex case will be reduced in Theorem 5 to the case of additive functions modulo 1. This reduction depends on the following theorem on u. a. p. arithmetical functions $h(n)$ of absolute value 1.

THEOREM 4. *If the function $h(n)$ is u. a. p. and satisfies $|h(n)| = 1$, $n = 1, 2, \dots$, then there exist an integer P , real numbers c_u and real-valued u. a. p. functions ψ_u , $u = 1, \dots, P$, such that*

$$h(u + nP) = \exp 2\pi i(c_u n + \psi_u(n));$$

$$(u = 1, \dots, P; n = 1, 2, \dots).$$

Theorem 4 is a special case of a known theorem concerning generalized almost periodic functions on groups.²

² E. R. van Kampen, *Journal of the London Mathematical Society*, vol. 12 (1937), pp. 3-6; the result needed is (2), as modified by the last remark on p. 4. Since this

Use will be made of the following lemmas:

I. If the function $\phi(n)$ satisfies for a fixed prime p the requirement

$$(11) \quad \phi(n) = \phi(p^k) \text{ if } p^k | n \text{ and } p^{k+1} \nmid n, \quad (k = 0, 1, 2, \dots),$$

then $\phi(n)$ is u. a. p. if and only if

$$(12) \quad \gamma = \lim \phi(p^k) \text{ exists,} \quad (k \rightarrow \infty).$$

Clearly this lemma³ is applicable both to the summands g_p of g and to the factors f_p of f .

Let ϕ_j denote, for $j = 1, 2, \dots$ the periodic functions defined by

$$\phi_j(n) = \phi(n) \text{ if } 1 \leq n \leq p^j; \quad \phi_j(n + p^j) = \phi_j(n) \text{ for every } n.$$

If (12) is satisfied, then $\phi_j(n) \rightarrow \phi(n)$ uniformly with respect to n as $j \rightarrow \infty$, so that $\phi(n)$ is u. a. p. In order to prove the converse, assume that $\phi(n)$ is u. a. p., and let $\mu = \mu(l)$ denote the translation function of $\phi(n)$, i. e., let $\mu(l)$ be for a fixed l the least upper bound of the function $|\phi(n + l) - \phi(n)|$ of $n (= 1, 2, \dots)$. It is clear from (11) that $\mu(l)$ is the least upper bound of the expression $|\phi(p^{k+j}) - \phi(p^k)|$ as $j = 1, 2, \dots$, where $k = k(l)$ denotes the number of times p occurs in the factorization of l . Since ϕ is u. a. p., the lower limit of $\mu(l)$ as $l \rightarrow \infty$ is 0. Thus the lower limit of $|\phi(p^{k+j}) - \phi(p^k)|$ as $k \rightarrow \infty$ is 0, so that (12) holds. This completes the proof of I.

II. The additive function $g(n)$ is bounded if and only if (7) holds, and in that case the series (5) representing $g(n)$ is uniformly convergent.

First, if (7) holds, then the absolute value of any $g(n)$ is not more than the sum of the series in (7), so that g is bounded. On the other hand, if $|g(n)| \leq M$ holds for every n , then $|\Sigma' g(p^k)| \leq M$, where the sum Σ' is taken over any finite number of distinct primes p and the exponents are arbitrary. But this implies that $\Sigma_p |g(p^k)|$ converges uniformly in the exponents k , and so (7) holds and (5) is uniformly convergent. This completes the proof of II.

modification has only been indicated, the following reduction of Theorem 4 to a conjecture of Wintner which was subsequently proved by Bohr (*Danske videnskaberens Selskab X*, vol. 10 (1930)) might be useful. Let P be a translation number of $h(n)$ which belongs to the value 1. Then $h_u(n) = h(u + nP)$ is, for $u = 1, \dots, P$, a u. a. p. function for which $|h_u(n + 1) - h_u(n)| \leq 1$. Thus the continuous u. a. p. function obtained from $h_u(n)$ by linear interpolation does not come arbitrarily near to 0. On applying to this continuous function the result of Bohr, one obtains the constant c_u and the function with the properties stated in Theorem 4.

³This lemma is closely related to a result of Toeplitz, *Mathematische Annalen*, vol. 98 (1927), p. 282.

III. If the additive function $g(n)$ is u. a. p., then (7) and (8) are satisfied. Since (7) follows from the boundedness of g , by II, it remains to prove that (8) holds.

Let p_0 be a fixed prime and let $g^0(n)$ denote the summand of $\overline{f}g$ in (5) which corresponds to p_0 . If $\mu(l)$ and $\mu^0(l)$ denote the translation functions of g and g^0 , it will first be shown that

$$(13) \quad \mu^0(2l) \leq \mu(2l) \text{ holds for } l = 1, 2, \dots$$

Let $\epsilon > 0$ be given and let a partial sum $g' = g'(n)$ of the uniformly convergent series (5) be determined in such a way that $|g(n) - g'(n)| < \epsilon$ for every n , and that the term of (5) which corresponds to the prime p_0 is one of the terms in g' . Then, if $\mu'(l)$ denotes the translation function of g' , one has $\mu'(2l) \leq \mu(2l) + 2\epsilon$. Since $\mu^0(2l)$ and $\mu(2l)$ are independent of ϵ , it is clear that (13) will follow if one proves that $\mu^0(2l) \leq \mu'(2l)$ for $l = 1, 2, \dots$.

To this effect, let l and n be given and let k be determined in such a way that neither n nor $n + 2l$ is divisible by p_0^k . Then, from (6),

$$g^0(n + rp_0^k) = g^0(n) \text{ and } g^0(n + 2l + rp_0^k) = g^0(n + 2l), \\ (r = 1, 2, \dots).$$

The number r may be determined in such a way that $n + rp_0^k$ and $n + 2l + rp_0^k$ are not divisible by any of the primes (except p_0) which correspond to summands occurring in f .⁴ For this r one has

$$g'(n + rp_0^k) = g^0(n + rp_0^k) \text{ and } g'(n + 2l + rp_0^k) = g^0(n + 2l + rp_0^k),$$

since every summand of g' (with the exception of g^0) is 0 at the values of n in question. Hence

$$|g'(n + 2l + rp_0^k) - g'(n + rp_0^k)| = |g^0(n + 2l) - g^0(n)|,$$

so that $\mu^0(2l) \leq \mu'(2l)$ by the definition of a translation function. This completes the proof of (13).

Since g is u. a. p., one has $\liminf \mu(2l) = 0$ as $l \rightarrow \infty$. Thus, from (13), also $\liminf \mu^0(2l) = 0$ as $l \rightarrow \infty$. But it is clear from the proof of I that the last relation implies the existence of the limit (8) for $p = p_0$. Since p_0 was an arbitrary prime number, the proof of III is complete.

The proof of Theorem 2 is now evident. In fact, if (7) and (8) hold for the additive function g , then g is, by II the sum of the uniformly con-

⁴ In fact, r is a solution of a system of linear congruences to distinct prime moduli. Note that at least one of the numbers n and $n + 2l + 1$ is even, so that the restriction to even translations $2l$ is necessary.

vergent series (5), the terms of which are u. a. p., by I. Thus g is a u. a. p. function. On the other hand, if the additive function g is u. a. p., then (7) and (8) hold, by III. Thus the proof of Theorem 2 is complete. And it is clear from I, II and Theorem 2, that the part of Theorem 1 which concerns additive functions is correct.

IV. A multiplicative function $f(n)$ satisfies (9) if and only if $f(n)$ is bounded and the product (2) representing $f(n)$ is uniformly convergent.

That (9) implies the boundedness and the uniform convergence of (2) is obvious. But, if (2) is uniformly convergent, then there exists for every $\epsilon > 0$ a prime number q such that $|1 - \Pi' f(p^k)| < \epsilon$ if the product Π' is taken over any finite number of primes p larger than q , and the k are arbitrary exponents. If in addition f is bounded, then the product $\Pi_p f(p^k)$ (taken over all primes) converges absolutely-uniformly with respect to the exponents k . Thus $\sum_p |f(p^k) - 1|$ converges uniformly with respect to the exponents k . And this, obviously, implies (9), so that the proof of IV is complete.

V. If $f(n)$ is a u. a. p. function and if (9) holds, i. e., if the product (2) is uniformly convergent, then (10) holds also.

The proof of this statement will be omitted, since it may be obtained from the proof of III by unessential modifications.

VI. If $f(n)$ is a non-negative, multiplicative u. a. p. function, then f satisfies condition (9).

Let M be an upper bound of f , so that $0 \leq f \leq M$ for every n . Then one has $0 \leq \Pi' f(p^k) \leq M$, where the product is taken over any finite number of distinct primes and the exponents k are arbitrary. Thus $\sum_p (1. u. b. f(p^k) - 1)$ is convergent, and (9) will follow if it is proved that $\sum_p (1 - \text{gr. l. b. } f(p^k))$ also is convergent (note that $f(p^0) = 1$).

Let the integer L be such that every group of L consecutive integers contains a translation number of $f(n)$ belonging to $\frac{1}{2}$. If $\sum_p (1 - \text{gr. l. b. } f(p^k))$ is divergent, so that $\Pi_p \text{gr. l. b. } f(p^k) = 0$, one can easily construct L numbers n_1, \dots, n_L such that

$$f(n_i) < \frac{1}{2M}; \quad (n_i, n_j) = 1 \text{ if } i \neq j; \quad (i, j = 1, \dots, L).$$

Since no two numbers n_i have a common factor, there exists an integer N such that

$$N + i \equiv n_i \pmod{n_i^2}, \text{ for } i = 1, \dots, L.$$

Then n_i and $(N + i)/n_i = n'_i$ are relative prime integers, so that

$$f(N+i) = f(n_i)f(n'_i) \leq f(n_i)M < \frac{1}{2}, \quad (i = 1, \dots, L).$$

This contradicts $f(1) = 1$, since there must exist at least one i_0 such that $0 < i_0 \leq L$ and $|f(N+i_0) - f(1)| < \frac{1}{2}$. Thus $\Sigma_p(1 - \text{gr. l. b.}_k(p^k))$ is convergent and so (9) holds for a non-negative, u. a. p., multiplicative function f . This completes the proof⁵ of VI.

It is clear from I and IV that the part of Theorem 1 which concerns multiplicative functions is an immediate consequence of Theorem 3. And, if a multiplicative function f satisfies (9) and (10), then the product (2) which represents f is uniformly convergent, by IV, and the factors of this product are u. a. p. functions, by I. Thus, in this case f is u. a. p., and the proof of the sufficiency of the conditions in Theorem 3 is complete. Now suppose that the real *non-negative* multiplicative function f is u. a. p. Then f satisfies (9), by VI, and hence f satisfies (10), by V. Thus the proof of the remaining part of Theorem 3 is complete in case the u. a. p. multiplicative function f is supposed to be non-negative.

It is clear from V, that the proof of Theorem 3 will be complete if the following analogue of VI for real-valued multiplicative functions is proved:

VII. *If f is a real-valued multiplicative u. a. p. function, then f satisfies (9).*

First, by VI, the function $|f|$ satisfies (9), so that

$$(14) \quad \Sigma_p \text{ l. u. b.}_k |f(p^k)| - 1 \text{ is convergent.}$$

Let p_1, \dots, p_m be a finite number of distinct primes, including all those primes for which $\text{gr. l. b.}_k |f(p^k)| = 0$. Then one has, as a consequence of (14),

$$\Pi_p \text{ gr. l. b.}_k |f(p^k)| > c, \quad 0 < c < 1,$$

where the product is taken over all primes except p_1, \dots, p_m .

Next, let $\lambda(n)$ be the multiplicative function which is defined by $\lambda(n) = 0$ or $\lambda(n) = f(n)/|f(n)|$ according as n is or is not divisible by at least one of the primes p_1, \dots, p_m . Then $\lambda(n)$ is a u. a. p. function. For if $\kappa(n)$ denotes the (periodic) function which is 1 or 0 according as n is or is not divisible by at least one of the primes p_1, \dots, p_m , then $\chi(n) = (1 - \kappa(n))|f(n)| + \kappa(n)$ is a real-valued u. a. p. function with the positive lower bound c . And so $\lambda(n) = (\chi(n)^{-1} - \kappa(n))f(n)$ also is a u. a. p. function.

Because $f(n)$ is real valued, the u. a. p. function $\lambda(n)$ can only assume the three values 0, 1 and -1 , so that $\lambda(n)$ is periodic. Since $\lambda(n)$ is multiplicative also, one has $\lambda(p^k) = 1$ for every k if the prime p is not a divisor of the primitive period of $\lambda(n)$. Hence it is clear from (14) and the definition of $\lambda(n)$ that (9) holds for $f(n)$. This completes the proof of VII, hence also the proof of Theorem 3.

⁵ This proof was communicated to me by P. Erdős.

It is easily seen that the proof of VII may be extended to the case where $\lambda(n)$ is restricted to assume a finite number of values. Thus one obtains:

VIII. *If the multiplicative function $f(n)$ is u. a. p. and if in addition there exists an integer r such that $f(p^k)^r$ is non-negative for every k unless p is one of a finite number of primes, then f satisfies conditions (9) and (10).*

In fact, if this finite number of primes is included among the primes p_1, \dots, p_m , of the proof of VII, then the function $\lambda(n)$ of that proof can only assume as values either 0 or $\exp(2\pi i k/r)$, $k = 1, \dots, r$. Thus $\lambda(n)$ will again be a periodic function, so that the proof may be completed as the proof of VII.

The reduction of the general multiplicative case to the case of additive functions modulo 1 will now be formulated and proved. A function $\psi(n)$ will be called additive modulo 1 if

$$\psi(n_1 n_2) \equiv \psi(n_1) + \psi(n_2) \pmod{1} \text{ if } (n_1, n_2) = 1$$

and n_1 and n_2 are not divisible by one of a finite number of primes p_1, \dots, p_w , while $\psi(n) = 0$ for all n which are divisible by one of these primes. Let $\{c\}$ denote the distance from c to the nearest integer.

IX. *The statement that conditions (9) and (10) are necessary for any multiplicative function to be u. a. p., is equivalent to the statement that the condition*

$$(15) \quad \Sigma_p \text{ l. u. b. } \{\psi(p^k)\} < +\infty$$

is necessary for a function $\psi(n)$ to be u. a. p., if $\psi(n)$ is additive modulo 1.

For a given u. a. p. multiplicative function $f(n)$, let the u. a. p. multiplicative function $\lambda(n)$ and the periodic function $\kappa(n)$ be defined as in the proof of VII. Then $h(n) = \lambda(n) + \kappa(n)$ is u. a. p. and satisfies $|h(n)| = 1$, so that Theorem 4 is applicable to $h(n)$. It may be assumed that the integer P of Theorem 4 is divisible by each of the primes p_1, \dots, p_m of the proof of VII. For certain values of the integer u of Theorem 4, one will have $h(u + nP) = \kappa(u + nP) = 1$, for every n . For the remaining values of u , including in particular $u = 1$, one has $\kappa(u + nP) = 0$ and

$$(16) \quad \lambda(u + nP) = \exp 2\pi i(c_u n + \psi_u(n)), \quad (n = 1, 2, \dots),$$

where c_u is a real constant and ψ_u a real-valued u. a. p. function. It will be shown that one may choose $c_u = 0$ for every u for which (16) holds. Since c_n is congruent modulo 1 to a u. a. p. function of n (either for all n or on any arithmetical sequence of values of n) if and only if c is rational, it will be sufficient to show that c_u is rational for every u for which (16) holds. Let such a u be fixed.

There exists an arithmetical sequence of values of n such that

$$(1 + P, u + nP) = 1 \quad \text{hence} \quad \lambda(1 + P)\lambda(u + nP) = \lambda((1 + P)(u + nP)).$$

Thus, since (16) holds for this u and for $u = 1$:

$$c_1 + \psi_1(1) + c_u n + \psi_u(n) \equiv (n + u + nP)c_u + \psi_u(n + u + nP) \pmod{1}$$

or

$$c_u Pn \equiv \psi_u(n + u + nP) - \psi_u(n) - \psi_1(1) - c_1 + uc_u \pmod{1}$$

for every n belonging to an arithmetical sequence. Since the right side is u. a. p. on this sequence, so is the left side. Thus c_u is rational, and one may suppose that $c_u = 0$ for every u for which (16) holds. The resulting functions $\psi_u(n)$ may be combined into one function ψ by placing $\psi(u + nP) = \psi_u(n)$. On applying the definition of $\lambda(n)$, one sees that

$$(17) \quad f(n)/|f(n)| = \exp 2\pi i \psi(n) \quad (n, p_1 p_2 \cdots p_m) = 1,$$

where $\psi(n)$ is a real-valued u. a. p. function, which may be defined to be 0 if n is divisible by one of the primes p_1, \dots, p_m . Since $f(n)$ is multiplicative, $\psi(n)$ is additive modulo 1. Thus if (15) were a necessary condition for a function $\psi(n)$ of this type to be u. a. p., then (15) would hold for $\psi(n)$, and it would be clear from (17), (14) and V that (9) and (10) were necessary conditions for any multiplicative f to be u. a. p. This proves one part of the assertion IX.

Now let $\psi(n)$ be u. a. p. and additive modulo 1. Then the function $f(n)$ which is defined by

$$\begin{aligned} f(n) &= \exp 2\pi i \psi(n) & \text{if } (n, p_1 p_2 \cdots p_m) = 1 \\ f(n) &= 0 & \text{if } (n, p_1 p_2 \cdots p_m) > 1 \end{aligned}$$

is u. a. p. and multiplicative. Thus if (9) were a necessary condition for a multiplicative function to be u. a. p., then this f would satisfy (9), so that $\psi(n)$ would satisfy (15). This completes the proof of IX.

As an example to Theorem 3, consider the function $f(n) = \sigma_\alpha(n)/n^\alpha$, $\alpha > 0$, where $\sigma_\alpha(n)$ denotes the sum of the α -th powers of the divisors of n . It will be shown that $\sigma_\alpha(n)/n^\alpha$ is uniformly almost periodic⁶ if and only if $\alpha > 1$. In fact, one has for this $f = f(n)$:

$$f(p^k) = \frac{p^{ak+a} - 1}{(p^a - 1)p^{ak}}, \quad f(p^k) - 1 = \frac{1 - p^{-ak}}{p^a - 1}$$

so that l. u. b. $|f(p^k) - 1| = (p^a - 1)^{-1}$, and (6) holds or does not hold according as $\alpha > 1$ or $\alpha \leq 1$. Since $f(p^k) \rightarrow (p^a - 1)^{-1} + 1$ as $k \rightarrow 0$ and since $f(n) > 0$ for every n , the above statement follows from Theorem 1.

THE JOHNS HOPKINS UNIVERSITY.

⁶ The function $\sigma_\alpha(n)/n^\alpha$ is almost periodic (B^λ) for every λ if $\alpha > 0$. This was shown *loc. cit.*¹, p. 625. Cf. Ramanujan, *Collected Works*, Cambridge (1927), p. 184, formula (6.1).

ADDITIVE FUNCTIONS AND ALMOST PERIODICITY (B^2).*

By PAUL ERDÖS and AUREL WINTNER.

1. By an additive function $f = f(n)$ is meant a sequence $f(1), f(2), \dots$, defined for every positive integer n in such a way that

$$(1) \quad f(n_1 n_2) = f(n_1) + f(n_2) \text{ whenever } (n_1, n_2) = 1; \quad (f(1) = 0).$$

Thus,

$$(2) \quad f(n) = \sum_{k=1}^{\infty} f^{(k)}(n) = \lim_{k \rightarrow \infty} f_k(n),$$

where $f_k = f_k(n)$ denotes, for fixed k , the additive function

$$(3) \quad f_k(n) = \sum_{j=1}^k f^{(j)}(n),$$

and $f^{(k)} = f^{(k)}(n)$ is the additive function which is defined in terms of the k -th prime number, p_k , as follows:

$$(4) \quad f^{(k)}(n) = \begin{cases} 0, & \text{if } n \not\equiv 0 \pmod{p_k}, \\ f(p_k^l), & \text{if } p_k^l | n \text{ and } p_k^{l+1} \nmid n, \end{cases}$$

($p_1 = 2, p_2 = 3, p_3 = 5, \dots$). Conversely, if $\{f(p_k^l)\}$ is any given double sequence of numbers, then (4), (3), (2) define $f^{(k)}, f_k, f$, respectively, as additive functions of n . In fact, all but a finite number of the terms of the infinite series (2) is zero for every fixed n .

The function $f(n)$ is called multiplicative if in condition (1) the sum $f(n_1) + f(n_2)$ becomes replaced by the product $f(n_1)f(n_2)$. Conditions which are either necessary or sufficient for the almost periodicity (B^2) of a multiplicative function $f(n)$ are implied by the results of a recent paper.¹ However, none of the results found *loc. cit.*¹ supplies a criterion which is necessary and at the same time sufficient for the almost periodicity (B^2) of a multiplicative function (not even if $f(n)$ is supposed to be real-valued). This situation is not surprising, since if a real-valued multiplicative function $f(n)$ changes its sign with the uniformity of statistical randomness (as does the Möbius function $f = \mu$), then the question as to a generalized almost periodic behavior

* Received December 4, 1939.

¹ E. R. van Kampen and Aurel Wintner, "On the almost periodic behavior of multiplicative number-theoretical functions," *American Journal of Mathematics*, vol. 62 (1940), pp. 613-626.

of $f(n)$ can involve problems of the same order of delicacy as do the relevant generalisations of the prime-number theorem, if not of the Riemann hypothesis. [While the prime-number theorem is equivalent to the statement that the n -average of $\mu(n) \exp(i\lambda n)$ exists for $\lambda = 0$, Davenport's results (*Quarterly Journal of Mathematics*, vol. 8 (1937), pp. 313-320), which were obtained by an application of the deep methods of Vinogradoff, imply that this average exists and vanishes for every real λ . In other words, all Fourier coefficients of $\mu(n)$ exist and vanish. Hence, $\mu(n)$ cannot be almost periodic (B). For if it were, the n -average of

$$|\mu(n) - (0 + 0 + \dots)| = |\mu(n)| = |\mu(n)|^2$$

ought to vanish. But this average is known to be $6\pi^{-2} \neq 0$.]

The object of the present paper is to show that the problem admits of a definitive solution in the case of additive, instead of multiplicative, functions. In fact, the question of almost periodicity (B^2) may then completely be answered by the following theorem:

An additive function $f = f(n)$ is almost periodic (B^2) if and only if both series

$$(i) \quad \sum_p \frac{f(p)}{p}; \quad (ii) \quad \sum_{l=1}^{\infty} \sum_p \frac{|f(p^l)|^2}{p^l}$$

are convergent.

This fact seems to be an arithmetical counterpart of a similar result concerning the case of linearly independent frequencies (cf. *loc. cit.*³, pp. 79-80). But we were unable to find the common source of these two parallel theorems.

It is understood that \sum_p denotes summation over all prime numbers, which are thought of as ordered according to magnitude (the series (i) need not be absolutely convergent).

2. If f' denotes the real, and f'' the imaginary, part of f , the function $f(n) = f'(n) + if''(n)$ is additive if and only if so are both functions $f'(n)$, $f''(n)$. Similarly, $f(n)$ is almost periodic (B^2) if and only if so are $f'(n)$ and $f''(n)$. Finally, it is clear from $|f|^2 = (f')^2 + (f'')^2$ that both series (i), (ii) are convergent if and only if so are the $2 + 2$ series which one obtains by writing f' and f'' for f in (i), (ii).

Consequently, it is sufficient to prove the italicized theorem for the case of real-valued additive functions. The possibility of this reduction is essential for the method to be applied. In fact, use will be made of a criterion which

recently² was proved to be necessary and sufficient for those real-valued additive functions which possess an asymptotic distribution function. Now, a generalization of this criterion for complex-valued functions is not known and seems to lead to essential difficulties.

The criterion in question states² that a real-valued additive function $f(n)$ has an asymptotic distribution if and only if both series

$$(I) \sum_p \frac{f^*(p)}{p}; \quad (II) \sum_p \frac{f^*(p)^2}{p}$$

are convergent, where $y^* = f^*(n)$ is defined, for $y = f(n)$, by placing

$$(5) \quad y^* = y \text{ or } y^* = 1 \text{ according as } |y| < 1 \text{ or } |y| \geq 1.$$

It follows that the convergence of both series (I), (II) is necessary for every (real-valued, additive) f which is almost periodic (B^2). In fact, it is known³ that almost periodicity in relative measure and so, in particular, almost periodicity in relative mean of any positive order ($=2$ in the present case) is always sufficient for the existence of an asymptotic distribution function.

2 bis. Suppose, in particular, that $f(p) = O(1)$ as $p \rightarrow \infty$. Then, since

$$(6) \quad \sum_{l=2}^{\infty} \sum_p \frac{1}{p^l} < \infty,$$

the series (ii) of § 1 is convergent if and only if so is the series

$$(7) \quad \sum_p \frac{f(p)^2}{p};$$

hence, one readily sees from (5) that the convergence of the series (i), (ii) which occur in the criterion of § 1 is equivalent to the convergence of the respective series (I), (II) which occur in the criterion of § 2.

3. For arbitrary additive functions f , the italicized statement of § 1 will be refined by exhibiting, in case of almost periodicity (B^2), a sequence of functions which are explicitly defined in terms of f , tend to f with reference

² Paul Erdős and Aurel Wintner, "Additive arithmetical functions and statistical independence," *American Journal of Mathematics*, vol. 61 (1939), pp. 713-721.

³ Børge Jessen and Aurel Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88, more particularly Theorem 24 (and Theorem 25).

to the metric of the space (B^2) , and are almost periodic (B^2) . In fact, it turns out that f cannot be almost periodic (B^2) unless it is almost periodic (B^2) in virtue of its expansion (2). In other words, if f is almost periodic (B^2) , then, on the one hand, each of the functions $f^{(1)}, f^{(2)}, \dots$ is almost periodic (B^2) , and, on the other hand,

$$(8) \quad M\{|f - f_k|^2\} \rightarrow 0 \text{ as } k \rightarrow \infty, \quad (f_k = \sum_{j=1}^k f^{(j)}),$$

where $M\{g\} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n g(l)$.

3 bis. Due to this fact, it will be possible to calculate the Fourier series of f in terms of the Ramanujan sums

$$(9) \quad c_m(n) = \sum_j' \exp(2\pi i \frac{j}{m} n), \text{ where } 1 \leq j \leq m \text{ and } (j, m) = 1.$$

In fact, the explicit form of the Fourier expansion of an arbitrary additive, almost periodic (B^2) function $f(n)$ turns out to be

$$(10) \quad f(n) \sim a_0 + \sum_l \sum_k a_{lk} c_{p_k^l}(n),$$

where $l = 1, 2, 3, \dots$, $k = 1, 2, 3, \dots$ and

$$(11) \quad a_0 = M\{f\}, \quad a_{lk} = \sum_{i=1}^{\infty} \frac{f(p_k^i) - f(p_k^{i-1})}{p^i}.$$

Since (9) consists of $\phi(m)$ terms (ϕ = Euler's function), and since $\phi(p^l) = p^l - p^{l-1}$, the Parseval relation belonging to (10) is

$$(12) \quad M\{|f|^2\} = |a_0|^2 + \sum_l \sum_k (p_k^l - p_k^{l-1}) |a_{lk}|^2.$$

4. It is easy to show that if f is such as to make the series (ii) of §1 convergent, then each of the functions f_k is almost periodic (B^2) .

To this end, use will be made of the following fact, proved *loc. cit.*¹ (Theorem II): If a function $g = g(n)$ of the positive integer n is such that, for some fixed prime number p , one has

$$(13) \quad g(n) = g(p^l) \text{ whenever } p^l | n \text{ and } p^{l+1} \nmid n,$$

then g is almost periodic (B^2) if and only if

$$(14) \quad \sum_{l=1}^{\infty} \frac{|g(p^l)|^2}{p^l} < \infty.$$

It is clear from (4) that condition (13) is satisfied by $g = f^{(k)}$ and

$p = p_k$, where k is arbitrarily fixed. Furthermore, if f is such as to make the series (ii) convergent, then, for every fixed k ,

$$(15) \quad \sum_{l=1}^{\infty} \frac{|f(p_k^l)|^2}{p^l} < \infty;$$

so that, since $f(p_k^l) = f^{(k)}(p_k^l)$ in view of (4), condition (14) also is satisfied by $g = f^{(k)}$ and $p = p_k$. Consequently, $f^{(k)}$ is almost periodic (B^2). Since k is arbitrary, and since the almost periodic (B^2) functions form a linear space, the almost periodicity (B^2) of f_k now follows from (3).

4 bis. It was shown *loc. cit.*¹ (Theorem III) that if a function $g(n)$ satisfies (13) for some fixed prime p and is almost periodic (B), then its Fourier expansion is

$$g(n) \sim M\{g\} + \sum_l a_l c_{p^l}(n), \text{ where } a_l = \sum_{i=l}^{\infty} \frac{g(p^i) - g(p^{i-1})}{p^i}.$$

It follows therefore from §4 that if f is such as to make the series (ii) convergent, then, for every k ,

$$f^{(k)}(n) \sim M\{f^{(k)}\} + \sum_l a_{lk} c_{p_k^l}(n), \text{ where } a_{lk} = \sum_{i=l}^{\infty} \frac{f^{(k)}(p_k^i) - f^{(k)}(p_k^{i-1})}{p^i}.$$

Hence, (10) with (11) will follow from (4) as soon as it is proved that, on the one hand, the convergence of the series (ii) is a necessary condition for the almost periodicity (B^2) of f , and that, on the other hand, f must satisfy (8) whenever it is almost periodic (B^2).

Proof of the sufficiency of the conditions.

From here on till the end of §9, the assumption will be that $f(n)$ is a real additive function for which both series (i), (ii) of §1 are convergent. The final result (§9) will be that $f(n)$ must then be almost periodic (B^2).

5. In terms of the given $f(n)$, define an $F(n)$ as follows: $F(n)$ is that additive function for which the double sequence $\{\{F(p_k^l)\}\}$ is given by

$$(16) \quad F(p^l) = \begin{cases} f(p^l), & \text{if } |f(p)| \geq 1; \\ f(p^l) - f(p), & \text{if } |f(p)| < 1, \end{cases}$$

where $p = p_k$ and $k = 1, 2, 3, \dots$

It is easy to see that the convergence of the series (ii) implies that

$$(17) \quad \sum_{l=1}^{\infty} \sum_p \frac{|F(p^l)|}{p^l} < \infty.$$

In fact, it is clear from (16) that the series (17) is majorized by $A + B + C$, where

$$A = \sum_p \frac{|F(p)|}{p}, \quad B = \sum_{l=2}^{\infty} \sum_p \frac{|f(p^l)|}{p^l}, \quad C = \sum_{l=2}^{\infty} \sum_p \frac{|f(p)|}{p^l},$$

and so it is sufficient to prove the convergence of these three series. But application of (16) to $l=1$ shows that $F(p) = 0$ unless $|F(p)| \geq 1$, in which case $F(p) = f(p)$; so that the series A reduces to

$$A = \sum_{|f(p)| \geq 1} \frac{|f(p)|}{p},$$

and is therefore convergent in virtue of the assumption that the series (ii) converges. It is clear from the same assumption and from (6), that also the series B is convergent. Finally, the series C may be written in the form

$$C = \sum_{l=2}^{\infty} \sum_{|f(p)| < 1} \frac{|f(p)|}{p^l} + \sum_{l=2}^{\infty} \sum_{|f(p)| \geq 1} \frac{|f(p)|}{p^l}.$$

But the convergence of the first of these two double series is assured by (6), while the second is, in view of

$$\sum_{l=2}^{\infty} \frac{1}{p^l} < \frac{1}{p}, \quad (p = 2, 3, 5, \dots),$$

majorized by

$$\sum_{|f(p)| \geq 1} \frac{|f(p)|}{p}.$$

Since the value of the latter series was seen to be $A < \infty$, the proof of (17) is now complete.

Similarly,

$$(18) \quad \sum_{l=1}^{\infty} \sum_p \frac{|F(p^l)|^2}{p^l} < \infty.$$

In fact, since $(a-b)^2 \leq 2(a^2 + b^2)$ for arbitrary real a, b , one sees from (16) that the series (18) is majorized by $A' + B' + C'$ where

$$A' = \sum_p \frac{|F(p)|^2}{p}, \quad C' = 2 \sum_{l=2}^{\infty} \sum_p \frac{|f(p^l)|^2}{p^l}, \quad B' = 2 \sum_{l=2}^{\infty} \sum_p \frac{|f(p)|^2}{p^l}.$$

And the proof for the convergence of these three series requires but a repetition (with obvious simplifications) of the above proof for the convergence of the three series A, B, C .

Notice that only the convergence of the second of the series (i), (ii) was used thus far. The same remark will hold for § 6.

6. It will now be shown that if $F_k(n)$ denotes the additive function

which belongs to the additive function $F(n)$ in the same way as (3) belongs to $f(n)$, then

$$(19) \quad \bar{M}\{|F - F_k|^2\} \rightarrow 0 \text{ as } k \rightarrow \infty,$$

where $\bar{M}\{|g|^2\} = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n |g(l)|^2$.

To this end, notice first that, by the definition (16) of the additive function $F(n)$,

$$\begin{aligned} \sum_{m=1}^n |F(m) - F_k(m)|^2 \\ \leq \sum_{l=1}^{\infty} \sum_{j=1}^{\infty} \sum_{p>k} \sum_{q>k} \frac{n}{p^l q^j} |F(p^l)F(q^j)| + \sum_{l=1}^{\infty} \sum_{p>k} \frac{n}{p^l} |F(p^l)|^2, \end{aligned}$$

where n and k are arbitrarily fixed, and the summation indices p, q run through those primes which exceed k . On writing this inequality in the form

$$\frac{1}{n} \sum_{m=1}^n |F(m) - F_k(m)|^2 \leq \left(\sum_{l=1}^{\infty} \sum_{p>k} \frac{|F(p^l)|}{p^l} \right)^2 + \sum_{l=1}^{\infty} \sum_{p>k} \frac{|F(p^l)|^2}{p^l},$$

keeping k fixed but letting $n \rightarrow \infty$, one sees that

$$(19 \text{ bis}) \quad \bar{M}\{|F - F_k|^2\} \leq \epsilon_k^2 + \epsilon'_k,$$

where

$$\epsilon_k = \sum_{l=1}^{\infty} \sum_{p>k} \frac{|F(p^l)|}{p^l}, \quad \epsilon'_k = \sum_{l=1}^{\infty} \sum_{p>k} \frac{|F(p^l)|^2}{p^l}.$$

But these sums ϵ_k, ϵ'_k are identical with the k -th remainders of the convergent series (17), (18), respectively, and tend therefore to zero as $k \rightarrow \infty$. Hence, (19) is implied by (19 bis).

7. If $G_k = G_k(n)$ denotes the additive function which belongs to the additive function

$$(20) \quad G = f - F$$

in the same way as f_k, F_k belong to f, F respectively, then obviously

$$(21) \quad G_k = f_k - F_k.$$

Thus, it is clear from (16) that, for any fixed k , the elements of the double sequence $\{\{G(p^l) - G_k(p^l)\}\}$ of the additive function $G(n) - G_k(n)$ of n are independent of l , i. e., that

$$(22) \quad G(p) - G_k(p) = G(p^2) - G_k(p^2) = G(p^3) - G_k(p^3) = \dots$$

for every prime p . It is also seen from (16) and (20) that

$$(23) \quad |G(p)| \leq 1$$

for every prime p .

Since the series (i) of § 1 is supposed to be convergent, it is clear from (20) and (17) that also

$$(24) \quad \sum_p \frac{G(p)}{p} \text{ is convergent.}$$

Similarly, since the series (ii) of § 1 is supposed to be convergent, it is clear from (20), from the Schwarz inequality

$$\sum_p \frac{|f(p)F(p)|}{p} \leq \left(\sum_p \frac{f(p)^2}{p} \right)^{\frac{1}{2}} \left(\sum_p \frac{F(p)^2}{p} \right)^{\frac{1}{2}},$$

and from (18), that

$$(25) \quad \sum_p \frac{G(p)^2}{p} < \infty.$$

8. Due to (22), it is now easy to transcribe the O -estimates applied *loc. cit.*² (p. 716) into o -estimates, which are to the effect that

$$(26) \quad \bar{M}\{|G - G_k|^2\} \rightarrow 0 \text{ as } k \rightarrow \infty.$$

In fact, (26) may be proved as follows:

If n and k are arbitrarily fixed, one readily verifies from (22) and from the definitions of the real additive functions G, G_k , that

$$\sum_{m=1}^n |G(m) - G_k(m)|^2 = \sum_{p > k} \sum'_{q > k} \left[\frac{n}{pq} \right] G(p)G(q) + \sum_{p > k} \left[\frac{n}{p} \right] G(p)^2,$$

where $[x]$ denotes the integral part of x , the prime of $\sum \sum'$ means that $p \neq q$, and the summation indices p, q run through those primes which exceed k (however, the sums on the right are finite sums for every fixed n , since

$$\left[\frac{n}{pq} \right] = 0 \text{ and } \left[\frac{n}{p} \right] = 0 \text{ whenever } pq > n \text{ and } p > n,$$

respectively). Consequently,

$$\begin{aligned} (26 \text{ bis}) \quad \frac{1}{n} \sum_{m=1}^n |G(m) - G_k(m)|^2 &\leq \left(\sum_{k < p \leq n} \frac{G(p)}{p} \right)^2 + 2 \sum_{n^{\frac{1}{2}} \leq p \leq n} \frac{G(p)}{p} \left(\sum_{k \leq q \leq n/p} \frac{G(q)}{q} \right) \\ &\quad + \frac{1}{n} \sum_{pq \leq n} |G(p)G(q)| + \sum_{p > k} \frac{G(p)^2}{p}. \end{aligned}$$

8 bis. As to the inner sum in the second of the four terms on the right

of (26 bis), one sees from (24) that if k is fixed and $\epsilon^{(k)}$ denotes the maximum of the function

$$\left| \sum_{k \leq q \leq n/p} \frac{G(q)}{q} \right|$$

of p and n on the range $n^{\frac{1}{2}} \leq p \leq n$; $n = 1, 2, \dots$, then $\epsilon^{(k)} \rightarrow 0$ as $k \rightarrow \infty$; while the absolute value of the whole second term on the right of (26 bis) has, for every n , the majorant

$$2 \sum_{n^{\frac{1}{2}} \leq p \leq n} \frac{|G(p)|}{p} \epsilon^{(k)} \leq 2\epsilon^{(k)} \sum_{n^{\frac{1}{2}} \leq p \leq n} \frac{1}{p},$$

by (23). Finally,

$$\frac{1}{n} \sum_{pq \leq n} |G(p)G(q)| \leq \frac{1}{n} \sum_{pq \leq n} 1, \text{ by (23).}$$

Thus, on keeping k fixed but letting $n \rightarrow \infty$, one sees from (26 bis) that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n |G(m) - G_k(m)|^2 \\ & \leq \limsup_{n \rightarrow \infty} \left\{ \left(\sum_{k < p \leq n^{\frac{1}{2}}} \frac{G(p)}{p} \right)^2 + 2\epsilon^{(k)} \sum_{n^{\frac{1}{2}} \leq p \leq n} \frac{1}{p} + \frac{1}{n} \sum_{pq \leq n} 1 \right\} + \sum_{p > k} \frac{G(p)^2}{p}. \end{aligned}$$

But p and q are prime numbers; so that

$$\sum_{n^{\frac{1}{2}} \leq p \leq n} \frac{1}{p} < \text{Const. and } \frac{1}{n} \sum_{pq \leq n} 1 \rightarrow 0$$

as $n \rightarrow \infty$. Hence,

$$\bar{M}\{|G - G_k|^2\} \leq \limsup_{n \rightarrow \infty} \left(\sum_{k < p \leq n^{\frac{1}{2}}} \frac{G(p)}{p} \right)^2 + \text{const.} \epsilon^{(k)} + \sum_{p > k} \frac{G(p)^2}{p}.$$

On letting here $k \rightarrow \infty$, and using the fact $\epsilon^{(k)} \rightarrow 0$ as $k \rightarrow \infty$, one sees from (24) and (25) that the proof of (26) is complete.

9. It is now easy to conclude that $f(n)$ is almost periodic (B^2) and satisfies (8).

In fact, since it was proved in § 4 that f_k is almost periodic (B^2) in virtue of the convergence of the series (ii), it is sufficient to show that

$$\bar{M}\{|f - f_k|^2\} \rightarrow 0 \text{ as } k \rightarrow \infty.$$

But the truth of this relation is implied by (19) and (26), since it is clear from (20) and (21) that

$$\bar{M}\{|f - f_k|^2\}^{\frac{1}{2}} \leq \bar{M}\{|F - F_k|^2\}^{\frac{1}{2}} + \bar{M}\{|G - G_k|^2\}^{\frac{1}{2}}.$$

Proof of the necessity of the conditions.

What remains to be proved is that the sufficient condition represented by the convergence of the two series (i), (ii) of § 1 is necessary as well. Thus, from now on till the end of the paper, the assumption will be that the $f(n)$ is any given real, additive function which is almost periodic (B^2).

10. Since $f(n)$ has an asymptotic distribution function, the two series (I), (II) of § 2 are convergent. And, in view of (5), the convergence of (II) implies that

$$(27) \quad \sum_{|f(p)| \geq 1} \frac{1}{p} < \infty.$$

In terms of the given $f(n)$, define an additive function $D(n)$ by placing

$$(28) \quad D = f - H,$$

where $H = H(n)$ denotes that additive function for which the double sequence $\{H(p^l)\}$ is given by

$$(29) \quad H(p^l) = \begin{cases} f(p^l), & \text{if } l \neq 1, \\ f(p), & \text{if } l = 1 \text{ and } |f(p)| \geq 1, \\ 0, & \text{if } l = 1 \text{ and } |f(p)| < 1, \end{cases}$$

($p = p_k$ and $k = 1, 2, 3, \dots$). Thus,

$$D(p^l) = \begin{cases} 0, & \text{if } l \neq 1, \\ 0, & \text{if } l = 1 \text{ and } |f(p)| \geq 1, \\ f(p), & \text{if } l = 1 \text{ and } |f(p)| < 1, \end{cases}$$

and so it is clear from (27) that one obtains two convergent series by writing D for f in (i)-(ii), § 1. Since the first half of the italicized statement of § 1 was already proved (§ 5-§ 9), it follows that $D(n)$ is almost periodic (B^2). Since $f(n)$ is almost periodic (B^2) by assumption, one sees from (28) that $H(n)$ is almost periodic (B^2).

In particular, $H(n)$ has a square-average

$$(30) \quad M\{H^2\} < +\infty.$$

11. In what follows, r will denote any of those prime numbers for which the absolute value of the given additive function f is not less than 1. Clearly, (27) may be written in the form

$$(31) \quad \prod_{|f(r)| \geq 1} \left(1 - \frac{1}{r}\right) > 0.$$

Since also the density of the *quadratifrei* integers is a positive number ($= 6\pi^{-2}$), a standard application of the sieve of Erathostenes shows that (31)

may be interpreted as follows: If n, j are positive integers and p is a prime, let $N = N(n, p, j)$ denote the number of those integers between 1 and n which are of the form $p^j s$, where s is *quadratifrei*, is not a multiple of p , and not a multiple of any of the primes r (defined by $|f(r)| \geq 1$). Then there exists a constant $\beta > 0$ which is independent of n, p, j and is such that

$$N = N(n, p, j) > \beta n p^{-j}.$$

Hence, it is clear from the definition (29) of the additive function $H(n)$, that

$$\sum_{m=1}^n H(m)^2 > \sum_{p^l < n} \frac{\beta n}{p^l} H(p^l)^2,$$

where the summation indices $p (= 2, 3, 5, \dots)$ and $l (= 1, 2, \dots)$ run through those of their combinations for which $p^l < n$. Thus, on writing this inequality in the form

$$\sum_{p^l < n} \frac{H(p^l)^2}{p^l} < \text{const.} \frac{1}{n} \sum_{m=1}^n H(m)^2, \quad (\text{const.} = \beta^{-1} < \infty),$$

and letting $n \rightarrow \infty$, one sees from (30) that

$$(32) \quad \sum_{l=1}^{\infty} \sum_p \frac{H(p^l)^2}{p^l} < \infty,$$

where p runs through all primes.

12. In view of (29), the content of (32) is that, on the one hand,

$$(33) \quad \sum_{l=2}^{\infty} \sum_p \frac{f(p^l)^2}{p^l} < \infty,$$

and, on the other hand,

$$(34) \quad \sum_{|f(p)| \geq 1} \frac{f(p)^2}{p} < \infty;$$

while (34) implies that

$$(35) \quad \sum_{|f(p)| \geq 1} \frac{|f(p)|}{p} < \infty.$$

Finally, as pointed out at the beginning of § 10 (cf. § 2), the series (I), (II) of § 2 are convergent. This means, in view of (5), that

$$(36) \quad \sum_{|f(p)| \leq 1} \frac{f(p)^2}{p} < \infty$$

and that also

$$(37) \quad \sum_{|f(p)| \leq 1} \frac{f(p)}{p} \text{ is convergent.}$$

Now, the convergence of the series (i) and (ii) of § 1 is clear from (37), (35) and (36), (34), (33), respectively.

STATISTICAL INDEPENDENCE AND STATISTICAL EQUILIBRIUM.*

By PHILIP HARTMAN and AUREL WINTNER.

Consider a conservative dynamical system which has a finite number of degrees of freedom and a Hamiltonian function possessing everywhere continuous partial derivatives of the second order. Suppose that some fixed value of the energy constant h determines a closed, bounded energy surface $\Omega = \Omega(h)$ in the phase-space; and that this Ω does not contain too many or too high critical points (e. g., that no point of Ω is an equilibrium solution of the system). If P is any point of Ω , the isoenergetic differential equations determine on Ω a unique phase-path P_t for which $P_0 = P$, and which exists for $-\infty < t < +\infty$. The resulting isoenergetic flow on Ω may also be described by placing $P_t = \tau_t P$, where τ_t , $-\infty < t < +\infty$, is for any fixed t a topological transformation of Ω into itself, and the function $\tau_t P$ of (t, P) is continuous on the product space of Ω and the t -axis. If one projects the euclidean Lebesgue measure of the phase-space on the energy surface Ω in the usual way,¹ and denotes by $\mu(E)$ the resulting Lebesgue measure of an arbitrary Borel subset E of Ω , then $\mu(\tau_t E) = \mu(E)$ for every E and t , since the isoenergetic differential equations which define τ_t satisfy the condition of Liouville.²

Since obviously $0 < \mu(\Omega) < \infty$, it may be assumed that $\mu(\Omega) = 1$. Thus, Birkhoff's ergodic theorem is applicable³ to the flow τ_t on Ω , and states that the path P_t has an asymptotic distribution function unless the initial condition $P = P_0$ is chosen on a set of μ -measure 0. It is understood that by the asymptotic distribution function ϕ_P of a path P_t is meant an absolutely additive set function $\phi_P = \phi_P(E)$, defined for all Borel subsets E of Ω and

* Received February 14, 1940.

¹ Cf. e. g., T. Levi-Civita, *Journal of Mathematics and Physics* (M. I. T.), vol. 13 (1934), pp. 22-23.

² For $n = 2$, cf. the explicit equations of G. D. Birkhoff, *Transactions of the American Mathematical Society*, vol. 18 (1917), pp. 211-212.

³ G. D. Birkhoff, *Proceedings of the National Academy*, vol. 17 (1931), pp. 656-660. The necessity of excluding possible discontinuity sets (cf. footnote ⁴) of the asymptotic distribution function belonging to a general P was pointed out by A. Wintner, *ibid.*, vol. 18 (1932), pp. 248-251; cf. P. Hartman and A. Wintner, *American Journal of Mathematics*, vol. 61 (1939), pp. 977-984.

Cf. also A. Wintner, *Nature*, vol. 145 (1940), pp. 225-226.

having the property that if E is any continuity set⁴ of ϕ_P , then the t -set defined by $P_t \subset E$ is relatively measurable, and has $\phi_P(E)$ as its relative measure.⁵ In other words, $\phi_P(E)$ is the probability that the path P_t , $-\infty < t < +\infty$, which is determined by $P = P_0$, be in the portion E of Ω . Since Ω is compact, the total probability $\phi_P(\Omega)$ is 1.

Any given Borel set E is a continuity set of $\phi_P(E)$ for almost all P . The proof of this fact will be omitted, since it readily follows from an estimate which occurs in Birkhoff's proof³ of the ergodic theorem.⁶

The fact just mentioned, when combined with Lebesgue's term-by-term integration of uniformly bounded sequences, obviously implies that

$$(1) \quad \mu(E) = \int_{\Omega} \phi_P(E) d_P \mu$$

for every Borel subset E of Ω .

Consider the product space $\Omega \times \Omega$ consisting⁷ of all pairs (P, Q) of points of Ω . Obviously, products $E \times F$ of Borel sets of Ω are Borel sets of $\Omega \times \Omega$. If on $\Omega \times \Omega$ a Lebesgue measure ν is defined by placing $\nu(E \times F) = \mu(E)\mu(F)$ for Borel sets $E \times F$, Birkhoff's ergodic theorem is obviously applicable to the product flow $\tau_t \times \tau_t$, with ν as the invariant measure on $\Omega \times \Omega$. Let ϕ_{PQ} denote the asymptotic distribution function of the path $(P_t, Q_t) = (\tau_t P, \tau_t Q)$, where it is understood that a set of initial points (P, Q) of ν -measure 0 must in general be excluded.

In what follows, use will be made of the fact that if $g_K(P)$ denotes the characteristic function of a Borel set K of Ω , then

$$(2) \quad \lim_{v-u \rightarrow \infty} \frac{1}{v-u} \int_u^v \left(\int_A g_E(P_t) d_P \mu \right) \left(\int_B g_F(Q_t) d_Q \mu \right) dt = \int_{A \times B} \phi_{PQ}(E \times F) d_{PQ} \nu$$

⁴ By a continuity set of a distribution function is meant any Borel set E which has the property that the distribution function attains the same value for the two Borel sets which represent the exterior and the closure of E . It is known that the Borel sets which are not continuity sets of a fixed distribution function are exceptional in the same sense as the discontinuity points of a fixed monotone function of a single variable.

⁵ A measurable set T of points of the t -axis is said to be relatively measurable if the Lebesgue t -measure of the common part of T and a finite interval $u \leq t \leq v$, when divided by the length $v-u$ of this interval, tends to a limit as $v-u \rightarrow \infty$; in which case this limit is called the relative measure of T .

⁶ As to this estimate, cf. N. Wiener, *Duke Mathematical Journal*, vol. 5 (1939), pp. 1-18 (cf. p. 2).

⁷ It should be emphasized that this product space is meant in the usual topological sense and is not, as it somehow became customary in ergodic theory, the symmetric product space. In other words, the points (P, Q) and (Q, P) of $\Omega \times \Omega$ will not be identified in the present paper.

holds for arbitrary Borel subsets A, B, E, F of Ω . In fact, if E and F are fixed, the remark which precedes (1) shows that $E \times F$ is a continuity set of ϕ_{PQ} for almost all points (P, Q) of $\Omega \times \Omega$. On the other hand, the ergodic theorem, when applied in its usual form to the fixed point function $f = f(P, Q) = g_E(P)g_F(Q)$ on $\Omega \times \Omega$, states that the limit

$$\lim_{v-u \rightarrow \infty} \frac{1}{v-u} \int_u^v g_E(P_t) g_F(Q_t) dt$$

exists for almost all points (P, Q) of $\Omega \times \Omega$. Since the definition of the asymptotic distribution function ϕ_{PQ} implies that the latter limit has the value $\phi_{PQ}(E \times F)$ whenever $E \times F$ is a continuity set of ϕ_{PQ} , it follows that, if E and F are fixed,

$$\lim_{v-u \rightarrow \infty} \frac{1}{v-u} \int_u^v g_E(P_t) g_F(Q_t) dt = \phi_{PQ}(E \times F)$$

holds for almost all points (P, Q) of $\Omega \times \Omega$. Hence, (2) follows by Lebesgue's theorem on term-by-term integration of uniformly bounded sequences.

Two solution paths P_t, Q_t on Ω are said to be statistically independent⁸ if the three asymptotic distribution functions $\phi_{PQ}, \phi_P, \phi_Q$ exist and satisfy the product condition

$$(3) \quad \phi_{PQ}(E \times F) = \phi_P(E) \phi_Q(F)$$

for all Borel sets $E \times F, E, F$ of $\Omega \times \Omega, \Omega, \Omega$ which are continuity sets of $\phi_{PQ}, \phi_P, \phi_Q$, respectively.

It turns out that the incompressible flows τ_t on Ω which possess this property of the statistical independence of almost all pairs of paths on Ω are interrelated with the incompressible flows τ_t on Ω which possess there a property of statistical equilibrium. From the physical point of view of statistical mechanics, this somewhat hidden interrelation between statistical independence and statistical equilibrium might perhaps have been expected. But we were unable to find any reference in the literature to the interrelation of these two physical concepts. On the other hand, the mathematical literature contains all of the tools necessary for this identification. In fact, Birkhoff's ergodic theorem, when stated as above in terms of asymptotic distribution functions,³ insures that the notion of statistical independence may be meant in its mathematical formulation, used loc. cit.⁷; while it is known that the notion of statistical equilibrium may be approached mathematically as follows:⁹

⁸ Cf. P. Hartman, E. R. van Kampen and A. Wintner, *American Journal of Mathematics*, vol. 61 (1939), pp. 477-486.

⁹ Cf. E. Hopf, *Proceedings of the National Academy*, vol. 18 (1932), pp. 333-340.

Suppose that the flow τ_t has the property that if there are given any Borel subset E of Ω and any "density of probability" as an integrable function $f = f(P)$ of P on Ω , then the probability carried by the set into which E is shifted by the flow τ_t tends to a limit as $t \rightarrow \infty$. If this condition is satisfied, i. e., if any given initial probability distribution is transformed in such a way as to become practically independent of t for large t , with reference to any Borel set E , then the flow τ_t is said to tend to statistical equilibrium. Since it may be shown¹⁰ that, instead of arbitrary integrable functions f , it is sufficient to consider characteristic functions $g_F(P)$ of arbitrary Borel sets F , the condition for a flow τ_t to tend to statistical equilibrium consists of the existence of the limit⁹

$$(4) \quad \lim_{t \rightarrow \infty} \mu(E_t \cdot F), \quad (E_t = \tau_t E),$$

for any pair E, F of Borel subsets of Ω . In fact, condition (4), where $A \cdot B$ denotes the common part of A and B , is precisely the previous definition, since obviously

$$(5) \quad \int_{E_t} g_F(P) d_P \mu = \mu(E_t \cdot F).$$

It is clear from (4) that if the limit (3) exists, its value is

$$(6) \quad \lim_{t \rightarrow \infty} \mu(E_t \cdot F) = \int_E \phi_F(F) d_P \mu \equiv \int_F \phi_F(E) d_P \mu,$$

where ϕ_F is the asymptotic distribution function of P_t . If it is only required that $\mu(E_t \cdot F)$ should become practically independent of t on the average, in the sense that, instead of the existence of the limit (6), one merely has⁹

$$(7) \quad \lim_{v-u \rightarrow \infty} \frac{1}{v-u} \int_u^v [\mu(E_t \cdot F) - \int_E \phi_F(F) d_P \mu]^2 dt = 0$$

for any pair E, F of Borel subsets of Ω , then the flow τ_t is said to tend to statistical equilibrium on the average. While it is clear that statistical equilibrium is sufficient for statistical equilibrium on the average, the converse is not true, at least¹¹ if the flow τ_t is not required to be one determined by an isoenergetic dynamical system. Incidentally, the content of the requirement (7) remains unchanged if one replaces the square $[]^2$ by $| [] |$; in fact, the integrand $[]^2$ is a bounded function of t , since $0 \leq \mu \leq 1$.

¹⁰ Cf. G. D. Birkhoff, *loc. cit.*²; N. Wiener, *loc. cit.*⁶.

¹¹ An example to this effect was given by B. O. Koopman and J. v. Neumann, *Proceedings of the National Academy*, vol. 18 (1932), pp. 255-263.

The interrelation between statistical independence and statistical equilibrium, as announced above, may now be formulated as follows: *In order that a flow be such as to make almost all pairs of paths statistically independent, it is sufficient (but, at least in case of a non-dynamical flow, not necessary) that it tend to statistical equilibrium; in fact, almost all pairs of paths are statistically independent if and only if the flow tends to statistical equilibrium on the average.*

That a flow τ_t which makes almost all pairs of paths statistically independent is necessarily a flow tending to statistical equilibrium on the average, is implied by the second half of Theorem 5 of E. Hopf.⁹ In fact, one can easily prove that his Theorem 5 is to the following effect: There is tendency toward statistical equilibrium on the average if and only if the condition (3), instead of being satisfied for all pairs (E, F) , is satisfied for symmetric pairs (E, E) only (it being understood that a zero set of pairs of points (P, Q) is always excluded). Apparently, it is this symmetry restriction¹² which has thus far disguised the interrelation between statistical independence and statistical equilibrium (either strict or average). In fact, as will be shown in the Appendix, two measurable functions of t need not be statistically independent if the condition corresponding to (3) is required for symmetric pairs $(E, F) = (E, E)$ only.

Nevertheless, it will now be shown that almost all pairs of paths are statistical independent in the case of a flow which tends to statistical equilibrium on the average.

To this end, suppose that the flow τ_t satisfies the average condition (7) for arbitrary Borel sets E, F , and write A, B for E, F ; so that

$$(7 \text{ bis}) \quad \lim_{v-u \rightarrow \infty} \frac{1}{v-u} \int_u^v [\mu(A_t \cdot B) - \int_A \phi_Q(B) d_Q \mu]^2 dt = 0.$$

Since both functions $\mu(A_t \cdot B), \mu(E_t \cdot F)$ of t lie between 0 and 1, it readily follows from (7) and (7 bis) that¹³

¹² Cf. footnote 7.

¹³ This depends on the following obvious remark: If $x(t), y(t)$ are bounded measurable functions for which there exist constants α, β such that

$$\frac{1}{v-u} \int_u^v [x(t) - \alpha]^2 dt \rightarrow 0, \quad \frac{1}{v-u} \int_u^v [y(t) - \beta]^2 dt \rightarrow 0,$$

then

$$\frac{1}{v-u} \int_u^v [x(t)y(t) - \alpha\beta]^2 dt \rightarrow 0,$$

as $v-u \rightarrow \infty$. It is sufficient to prove this in case $x(t)$ and $y(t)$ represent the same

$$\lim_{v \rightarrow \infty} \frac{1}{v-u} \int_u^v [\mu(A_t \cdot B) \mu(E_t \cdot F) - (\int_A \phi_P(B) d_P \mu) (\int_E \phi_Q(F) d_Q \mu)]^2 dt = 0,$$

or, if B and E are interchanged,

$$\lim_{v \rightarrow \infty} \frac{1}{v-u} \int_u^v [\mu(A_t \cdot E) \mu(B_t \cdot F) - (\int_A \phi_P(B) d_P \mu) (\int_B \phi_Q(F) d_Q \mu)]^2 dt = 0.$$

On the other hand, the identities (2) and (5) imply that

$$\int_{A \times B} \phi_{PQ}(E \times F) d_{PQ} \nu = \lim_{v \rightarrow \infty} \frac{1}{v-u} \int_u^v \mu(A_t \cdot E) \mu(B_t \cdot F) dt.$$

But comparison of the last two relations gives

$$\int_{A \times B} \phi_{PQ}(E \times F) d_{PQ} \nu = (\int_A \phi_P(E) d_P \mu) (\int_B \phi_Q(F) d_Q \mu).$$

Hence, by Fubini's theorem,

$$\int_{A \times B} \phi_P(E \times F) d_{PQ} \nu = \int_{A \times B} \phi_P(E) \phi_Q(F) d_{PQ} \nu,$$

since $\nu = \nu(E \times F)$ was defined as the product measure $\mu(E) \mu(F)$. Since the factors A, B of the integration domain $A \times B$ are arbitrary Borel sets of Ω , it follows from the separability of Ω , that the condition (3) of statistical independence is satisfied by almost all points (P, Q) of $\Omega \times \Omega$.

* * * * *

If, instead of two paths P_t, Q_t , one considers n paths P_t, Q_t, \dots, R_t , their statistical independence is defined by the requirement^s

$$(3 \text{ bis}) \quad \phi_{PQ \dots R}(E \times F \times \dots \times G) = \phi_P(E) \phi_Q(F) \dots \phi_R(G),$$

which reduces for $n=2$ to (3). It is known^s that $n=3$ given functions

function, and then successively choose the latter function to be $x(t), y(t), x(t) + y(t)$. But if $x(t)$ is bounded, then

$$[x(t)^2 - \alpha^2]^2 = [x(t) + \alpha]^2 [x(t) - \alpha]^2 \leq \text{const.} [x(t) - \alpha]^2.$$

(It is seen that it is sufficient to assume the boundedness of only one of the two functions x, y .)

need not be statistically independent if any of the three pairs which may be selected from them consists of two statistically independent functions. This might be one of the reasons why, on the basis of mere time averages of the solutions of the (isoenergetic) differential equations of classical dynamics, no mathematical theory has been developed thus far for physical facts of the type of the Maxwell-Boltzmann distribution or of the H -theorem. For these facts are asymptotic statements of the same type as is the validity of the normal distribution law in theory of errors; so that the number n of *independent* realizations of one and same model must be chosen arbitrarily large, since no statement can be made for a fixed n (in particular, for $n = 2$).

But it turns out that, due to the fact that the product spaces considered are *not* the symmetric product spaces,⁷ it is not difficult to pass from $n = 2$ to any n . While this sounds surprising in view of the examples just mentioned,⁸ all that actually happens is that n -uples of paths, possibly exceptional from the point of view of independence, are contained in zero sets which may vary with n . In other words, *if the flow makes the two paths P_t, Q_t statistically independent for almost all choices of (P, Q) on $\Omega \times \Omega$, then it also makes the n paths P_t, Q_t, \dots, R_t statistically independent for almost all choices (P, Q, \dots, R) on $\Omega \times \Omega \times \dots \times \Omega$, where n is arbitrary and it is understood that the sets excluded are zero sets with reference to the product measures (of μ) on $\Omega \times \Omega$ and $\Omega \times \Omega \times \dots \times \Omega$, respectively.*

In fact, it is clear that the calculation following (7 bis) may be carried out so as to show that the assumption (7) implies the statistical independence of almost all n -uples of solution paths, not only for $n = 2$ but for arbitrary n . Hence, the italicized statement follows from the fact that the requirement (7) of ultimate statistical equilibrium on the average was seen to be equivalent to the requirement of the statistical independence of almost all pairs of paths.

It follows that *the flow satisfies the requirement (7) of ultimate statistical equilibrium on the average if and only if the paths in almost all n -uples of paths are statistically independent.* This fact is, from the physical point of view, more important than the equivalent criterion in which n is restricted to $n = 2$. In fact, it now becomes admissible to consider a product space $\Omega \times \Omega \times \dots \times \Omega$ of arbitrarily many factors, thus introducing the flow on Ω in n independent copies, and then make $n \rightarrow \infty$. But this is precisely the relevant mathematical assumption of the theory of limit distributions in statistical mechanics.

If the incompressible flow τ_t on Ω , instead of satisfying any statistical assumption, is such as to make the asymptotic distribution function ϕ_P of the path P_t independent of the initial condition P (for almost all P), then the flow τ_t is necessarily metrically transitive, since (1) then reduces to

$\phi_P = \mu$ for almost all P . In particular, the class of those flows which satisfy the requirement (6) of ultimate statistical equilibrium and are at the same time such as to make ϕ_P independent of P for almost all P , is identical with the class of the flows to which the (in some respect misleading) name "mixture" was given.

Hedlund¹⁴ has recently proved that if Ω is a two-dimensional Riemannian manifold of constant negative curvature, of finite connectivity and of finite area, then the geodesic flow on Ω is a mixture. It follows, therefore, from the last italicized theorem, that the geodesic flow on any such Ω makes the paths of almost all n -uples of geodesics statistically independent of each other. Notice that in this example one has, besides statistical independence, asymptotic equidistribution of almost all paths; so that no example of an isoenergetic dynamical system seems to be known in which almost all pairs of paths are statistically independent but which is not metrically transitive.

APPENDIX.

It is known⁸ that two real measurable functions $x(t), y(t), -\infty < t < +\infty$, are statistically independent if and only if the Fourier average

$$(I) \quad \Lambda(u, v) = \lim_{s-r \rightarrow \infty} \frac{1}{s-r} \int_r^s \exp i\{ux(t) + vy(t)\} dt$$

exists uniformly in every fixed bounded domain of a real (u, v) -plane and satisfies the functional equation

$$(II) \quad \Lambda(u, v) = \Lambda(u, 0)\Lambda(0, v).$$

On the other hand, if instead of statistical independence, which corresponds to (3), one requires that the condition corresponding to (3) be satisfied for symmetric pairs $(E, F) = (E, E)$ only, then an obvious adaptation of the considerations applied loc. cit.⁸ shows that (II) must be replaced by the weaker condition

$$(III) \quad \Lambda(u, v) + \Lambda(v, u) = \Lambda(u, 0)\Lambda(0, v) + \Lambda(v, 0)\Lambda(0, u),$$

[which is again necessary and sufficient, provided that (I) exists uniformly in every fixed bounded domain of the (u, v) -plane, i. e., provided that the vector $(x(t), y(t))$ has an asymptotic distribution function]. But it is easy to construct a pair $(x(t), y(t))$ which satisfies (III) without satisfying (II). Actually, the pair will be chosen periodic in t ; so that (I) reduces to

¹⁴ G. A. Hedlund, *Annals of Mathematics*, vol. 40 (1939), pp. 370-383.

$$(IV) \quad \Lambda(u, v) = \int_0^1 \exp i\{ux(t) + vy(t)\} dt,$$

if the period is 1.

First, define a function L of two real variables u, v by placing

$$9L(u, v) = 1 + e^{i(u+v)} + e^{2i(u+v)} + 2e^{iu} + 2e^{2iv} + 2e^{i(2u+v)}.$$

Then an easy calculation shows that the functional equation (III) is, and that (II) is not, satisfied by $\Lambda = L$.

On the other hand, the function $L(u, v)$ is a trigonometric polynomial in which the coefficients of the exponentials are positive and have the sum 1. This means that $L(u, v)$ is the Fourier-Stieltjes transform of a 2-dimensional purely discontinuous distribution function (with a finite number of jumps). Hence, it is clear that one can choose on the interval $0 \leq t \leq 1$ two step functions $x(t), y(t)$ for which $\Lambda = L$ satisfies (IV).

Incidentally, the trigonometric polynomial $L(u, v)$ is seen to satisfy the symmetry relation $L(u, 0) = L(0, u)$. This means that the two functions $x(t), y(t)$ have the same distribution function.

QUEENS COLLEGE,
THE JOHNS HOPKINS UNIVERSITY.

ON AN ASYMPTOTIC FORMULA FOR THE FOURIER TRANSFORMS OF DISTRIBUTIONS ON CERTAIN CURVES.*

By E. K. HAVILAND.

The smoothness of infinite convolutions of the type occurring in the theory of the Riemann zeta-function has been treated by an estimate of Fourier-Stieltjes transforms of the distributions on convex curves. An earlier method¹ of obtaining such an estimate consisted in an extension of the usual estimate of the Bessel functions J_n , making use of a lemma of van der Corput and an assumption that the spectra are sufficiently smooth convex curves. The resulting estimate has then been refined² in such a way as to yield an asymptotic formula also. In the case where merely an appraisal is desired, the foregoing method has been superseded by a simpler and more general one,³ quite elementary in nature, which is free of the restrictions of dimensionality, analyticity and convexity imposed by the earlier method. This latter method does not, however, admit of obtaining an asymptotic formula, and it is the purpose of the present paper to obtain such a formula without the restriction of convexity and with fewer restrictions on the smoothness of the curves. The increased generality is obtained largely by following a method of Hartman⁴ for obtaining an asymptotic formula for exponential integrals.

Let $x = x(\phi)$, $y = y(\phi)$, where $0 \leq \phi < 2\pi$, be a parametric representation of a Jordan curve, S , to be described more precisely below, in the (x, y) -plane. Let $\sigma = \sigma(E)$ be an absolutely additive set function defined, for every Borel set, E , of the (x, y) -plane, by setting $\sigma(F)$ equal to $1/2\pi$ times the linear measure of those ϕ for which $(x(\phi), y(\phi))$ is contained in FS , if F is any open set in the plane. In particular, it is seen that S is the

* Received November 22, 1939.

¹ Cf. A. Wintner, "Upon a statistical method in the theory of diophantine approximations," *American Journal of Mathematics*, vol. 55 (1933), pp. 309-331; B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta-function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

² Cf. E. K. Haviland and A. Wintner, "On the Fourier transforms of distributions on convex curves," *Duke Mathematical Journal*, vol. 2 (1936), pp. 712-721.

³ Cf. A. Wintner, "On the smoothness of infinite convolutions of the type occurring in the theory of the Riemann zeta-function," *American Journal of Mathematics*, vol. 61 (1939), pp. 231-236.

⁴ Cf. Philip Hartman, "An asymptotic formula for exponential integrals," *American Journal of Mathematics*, vol. 62 (1940), pp. 115-121.

spectrum⁵ of σ . From the definition of Lebesgue and of Radon integrals, it follows that

$$(1) \quad \Lambda(u, v; \sigma) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp[i(ux + vy)] d_{xy} \sigma(E) \\ = \frac{1}{2\pi} \int_0^{2\pi} \exp[i(ux(\phi) + vy(\phi))] d\phi.$$

On setting $u = r \cos \psi$ and $v = r \sin \psi$, one obtains

$$(2) \quad \Lambda = \Lambda(r \cos \psi, r \sin \psi; \sigma) = \frac{1}{2\pi} \int_0^{2\pi} \exp[irh(\phi; \psi)] d\phi,$$

where

$$(2a) \quad h(\phi, \psi) = x(\phi) \cos \psi + y(\phi) \sin \psi.$$

It will be assumed that

- (i) $x(\phi)$ and $y(\phi)$ possess second derivatives of bounded variation;
- (ii) $h'(\phi; \psi)$ has, for any fixed ψ , exactly n zeros on the curve S and these zeros are all simple. Furthermore the zeros of $h''(\phi; \psi)$ are all simple⁶ and n in number. Here and in what follows n is a fixed positive integer and a prime denotes partial differentiation with respect to ϕ . As a consequence of (i), $h''(\phi) = x''(\phi) \cos \psi + y''(\phi) \sin \psi$ is continuous on the torus T : ($0 \leq \phi < 2\pi$; $0 \leq \psi < 2\pi$). As a consequence of (ii), the zeros of h'' separate those of h' . Thus the convex curves previously treated² are included as a proper subclass of the curves S now considered.

Under the foregoing assumptions, it will be shown that

$$(3) \quad \Lambda = (2\pi r)^{-\frac{1}{2}} \sum_{k=1}^n \{ [h''(\phi_{4k-1}(\psi); \psi)]^{-\frac{1}{2}} \exp[i(rh(\phi_{4k-1}(\psi); \psi) + \pi/4 + \pi/4)] \\ + [-h''(\phi_{4k-3}(\psi); \psi)]^{-\frac{1}{2}} \exp[i(rh(\phi_{4k-3}(\psi); \psi) - \pi/4)] \} + o(r^{-\frac{1}{2}}),$$

where the o -term holds uniformly for all ψ , and

$$\phi_{4k-3} = \phi_{4k-3}(\psi), \quad \phi_{4k-1} = \phi_{4k-1}(\psi), \quad (k = 1, \dots, n/2),$$

represent the zeros of h' on S , the former corresponding to maxima of h and the latter to minima. It will be observed that the $o(r^{-\frac{1}{2}})$ of the present paper replaces the $O(r^{-1})$ of the previous treatment, so that we now get precisely an asymptotic formula without a remainder term.

The proof of (3) proceeds as follows. First, the minimum distance between a zero of h' and a zero of h'' has, for reasons of continuity, a positive

⁵ For the definition of the spectrum, cf. A. Wintner, "On the addition of independent distributions," *American Journal of Mathematics*, vol. 56 (1934), pp. 8-16.

⁶ (ii) might be generalized, under suitable assumptions, to the case where the second derivative has multiple zeros.

lower bound $\bar{\xi}$ independent of ψ . Let $\phi_{2k} = \phi_{2k}(\psi)$, ($k = 1, \dots, n$), be the zeros of $h''(\phi; \psi)$ and let them be so situated that

$$\phi_1 < \phi_2 < \phi_3 < \phi_4 < \dots < \phi_{2n-3} < \phi_{2n-2} < \phi_{2n-1} < \phi_{2n} < \phi_1 + 2\pi.$$

Finally, let $2n$ numbers η_k be so chosen, as indicated more precisely below, that

$$\phi_1 < \eta_1 < \phi_2 < \eta_2 < \phi_3 < \dots < \phi_{2n} < \eta_{2n} < \phi_1 + 2\pi.$$

As $h''(\phi; \psi)$ is continuous on the torus T , it is clear that, if we write

$$h''(\phi; \psi) = h''(\phi_k; \psi) + \omega(\phi_k; \phi; \psi), \quad (k = 1, 2, \dots, 2n),$$

then $\omega(\phi_k; \phi; \psi)$ will possess the same property, so that to a preassigned $\epsilon > 0$ there corresponds a $\delta = \delta(\epsilon)$, independent of ϕ_k and of ψ , such that

$$|\omega(\phi_k; \phi; \psi)| < \epsilon$$

for all ϕ such that $|\phi - \phi_k| < \delta$.

Moreover, it is clear from (ii) that there exists a positive constant α such that $|h''(\phi_1(\psi); \psi)| > \alpha$ for every ψ . Then one may choose $\eta_1 = \eta_1(\psi)$ so that it lies between $\phi_1 + \xi/3$ and $\phi_1 + 2\xi/3$ for all ψ , where $\xi = \min(\bar{\xi}, \delta(\alpha/4))$ and is therefore independent of ψ , and a similar choice will be made for the remaining η_k 's.

From (2),

$$(4) \quad \Lambda = \frac{1}{2\pi} \left\{ \int_{\phi_1}^{\eta_1} + \int_{\eta_1}^{\phi_2} + \int_{\phi_2}^{\eta_2} + \dots + \int_{\eta_{2n-1}}^{\phi_{2n}} + \int_{\phi_{2n}}^{\phi_1 + 2\pi} \right\} \\ \equiv J_1 + J_2 + J_3 + \dots + J_{3n-1} + J_{3n},$$

say.

These integrals fall essentially into two classes: those, such as J_1 , in which the integration range possesses an end point ϕ_{2k-1} , ($k = 1, 2, \dots, n$), and those, such as J_2 , with an integration range containing a point ϕ_{2k} , ($k = 1, 2, \dots, n$), in its interior. In order to treat J_1 , set for $\phi_1 \leq \phi \leq \eta_1$, where $\phi_1 = \phi_1(\psi)$ and $\eta_1 = \eta_1(\psi)$,

$$(5) \quad t^2 = h(\phi_1; \psi) - h(\phi; \psi)$$

for every fixed ψ , corresponding to the fact that ϕ_1 is a simple zero of h by assumption (ii). On taking the positive square root,

$$(6) \quad t = |h(\phi_1; \psi) - h(\phi; \psi)|^{\frac{1}{2}},$$

so that as ϕ increases steadily from ϕ_1 to η_1 , the variable t increases steadily from zero to a quantity $a_1(\psi) = |h(\phi_1(\psi); \psi) - h(\eta_1(\psi); \psi)|^{\frac{1}{2}}$, which has a positive lower bound β independent of ψ in virtue of (ii). Hence t is in $[\phi_1(\psi), \eta_1(\psi)]$ a monotone, continuous function of ϕ .

Moreover, if a dot represents partial differentiation with respect to t ,

$$(7) \quad \dot{\phi} = -2t/h'(\phi(t, \psi); \psi), \text{ if } 0 < t \leq a_1(\psi),$$

so

$$(8) \quad J_1 = -\frac{1}{\pi} \exp[i\pi h(\phi_1(\psi); \psi)] \int_0^{a_1(\psi)} \exp[-i\pi t^2] t/h'(\phi(t, \psi); \psi) dt.$$

The integral in (8) is of the form

$$(9) \quad \int_0^{a_1(\psi)} f(t) \exp[-i\pi t^2] dt,$$

where

$$(10) \quad f(t) = f(t; \psi) = t/h'(\phi(t, \psi); \psi).$$

It is known⁴ that the integral (9) can, for every fixed ψ be evaluated asymptotically under the assumption that $f(t) = f(t; \psi)$ is of bounded variation in $[0, a_1(\psi)]$. That the function (10) possesses this property may be seen as follows:

Applying Taylor's Theorem with the integral form of the remainder, one obtains

$$(11) \quad h(\phi; \psi) = h(\phi) = h(\phi_1) + (\phi - \phi_1)h'(\phi_1) + \int_0^{\phi - \phi_1} (\phi - \phi_1 - s)h''(\phi_1 + s)ds,$$

where $\phi_1 = \phi_1(\psi)$. Since $h'(\phi_1) = 0$, (6) becomes, after a change of integration variable in (11),

$$(12) \quad t = \left\{ - \int_{\phi_1}^{\phi} (\phi - s)h''(s)ds \right\}^{\frac{1}{2}}.$$

Similarly, we have

$$(13) \quad h'(\phi; \psi) = h'(\phi) = \int_{\phi_1}^{\phi} h''(s)ds.$$

Since by hypothesis $h''(\phi; \psi) = h''(\phi)$ is, for all fixed ψ , a continuous function of ϕ and since $h''(\phi_1) \neq 0$, we may write, as above,

$$h''(\phi) = h''(\phi_1) + \omega(\phi); \quad |\omega(\phi)| < \epsilon \text{ if } |\phi - \phi_1| < \delta(\epsilon).$$

Then (12) becomes

$$t = \left\{ -\frac{1}{2}(\phi - \phi_1)^2 h''(\phi_1) - \int_{\phi_1}^{\phi} (\phi - s)\omega(s)ds \right\}^{\frac{1}{2}}$$

and (13) becomes

$$h'(\phi) = (\phi - \phi_1)h''(\phi_1) + \int_{\phi_1}^{\phi} \omega(s)ds.$$

Substituting these values into (10), we obtain

$$(14) \quad f(t) = f(t; \psi) = \frac{\left\{ -\frac{1}{2}h''(\phi_1(\psi); \psi) - (\phi - \phi_1)^{-2} \int_{\phi_1}^{\phi} (\phi - s)\omega(s)ds \right\}^{\frac{1}{2}}}{h''(\phi_1(\psi); \psi) + (\phi - \phi_1)^{-1} \int_{\phi_1}^{\phi} \omega(s)ds}$$

$$(15) \quad = \{-\frac{1}{2}h_1'' - L_{11}\}^{\frac{1}{2}} / \{h_1'' + L_{10}\},$$

where

$$(16) \quad L_{jp} = (\phi - \phi_j)^{-p-1} \int_{\phi_j}^{\phi} (\phi - s)^p \omega(s) ds.$$

Now the L_{1p} , ($p = 0, 1$), are, for fixed ψ , continuous functions of ϕ for $\phi_1(\psi) < \phi \leq \eta_1(\psi)$, and

$$(17) \quad |L_{jp}| \leq (\phi - \phi_j)^{-1} \int_{\phi_j}^{\phi} |\omega(s)| ds < \epsilon, \text{ if } 0 < |\phi - \phi_j| < \delta(\epsilon),$$

so that the L_{1p} are, for fixed ψ , continuous functions of ϕ in $\phi_1(\psi) \leq \phi \leq \eta_1(\psi)$ also if we define $L_{jp} = 0$ for $\phi = \phi_j$. Furthermore, as pointed out above, $\delta = \delta(\epsilon)$ in (17) is independent of ψ and of ϕ_j on T . By virtue of the choice of η_1 and of the existence of the quantity⁷ α , it follows that for all ϕ , ($\phi_1 \leq \phi \leq \eta_1$), and for all ψ ,

$$(18) \quad |h_1'' + L_{10}| > \alpha/2 \text{ and } |-\frac{1}{2}h_1'' - L_{11}| > \alpha/4.$$

From (15), (16), (17) and (18), it is clear that f is a continuous function of ϕ in $\phi_1 \leq \phi \leq \eta_1$ and since ϕ is, for fixed ψ , a continuous function of t in $0 \leq t \leq a_1(\psi)$, it is seen that $f(t; \psi)$ possesses the same property. Then as $\dot{\phi} = -2f$, it is clear that $\dot{\phi}$ tends to a definite limit as $t \rightarrow +0$, which implies that $\dot{\phi}$ exists and (7) holds at $t = 0$ also.

We now proceed to show that $f(t) = f(t; \psi)$, as defined by (10), is a function of bounded variation in t for $0 \leq t \leq a_1(\psi)$. In the first place, if a function $f(\phi)$ is of bounded variation in ϕ and ϕ , in turn, is a monotone continuous function of t , say in $[0, a_1]$, then $f(\phi(t))$ is a function of bounded variation in t in an appropriate interval. Consequently, by virtue of the remark following equation (6), we need prove only that f , as a function of ϕ , is of bounded variation. This we do by using (15) and the following familiar properties of functions of bounded variation:

- (α) The product of two functions of bounded variation in an interval $[a, b]$ is a function of bounded variation there.
- (β) If a function $F(x)$ is of bounded variation in $[a, b]$ and if $F(x) > \gamma > 0$ there, then $(F(x))^{\frac{1}{2}}$ and $(F(x))^{-1}$ are of bounded variation there.
- (γ) The product of two positive monotone non-decreasing (non-increasing) functions is again a positive monotone non-decreasing (non-increasing) function.
- (δ) $a(x - b) + c$ is monotone increasing (decreasing) if $a > 0$ ($a < 0$).
- (ϵ) The mean value over a finite interval of a function of bounded variation

⁷ Introduced prior to equation (4).

(monotone function) is again a function of bounded variation (monotone function).⁸

We now apply the foregoing results to the function $F(\phi)$ defined by the right-hand member of (14), considering first the second term in the numerator, which may be written

$$(19) \quad (\phi - \phi_1)^{-1} \int_{\phi_1}^{\phi} \omega(s) \frac{\phi - s}{\phi - \phi_1} ds = (\phi - \phi_1)^{-1} \int_{\phi_1}^{\phi} \omega(s) \left[1 - \frac{s - \phi_1}{\phi - \phi_1} \right] ds.$$

By hypothesis, $\omega(s) = \omega_1(s) - \omega_2(s)$, where ω_1, ω_2 are monotone non-decreasing. Since ω_1, ω_2 are both bounded, there exists a positive constant, C , such that $\omega_1(s) + C, \omega_2(s) + C$ are positive for all s in $[\phi_1, \eta_1]$. Then the left-hand member of (19) may be written in the form

$$\begin{aligned} (20) \quad & (\phi - \phi_1)^{-1} \int_{\phi_1}^{\phi} (\omega_1(s) + C) ds \\ & - (\phi - \phi_1)^{-1} \cdot (\phi - \phi_1)^{-1} \int_{\phi_1}^{\phi} (\omega_1(s) + C)(s - \phi_1) ds \\ & - (\phi - \phi_1)^{-1} \int_{\phi_1}^{\phi} (\omega_2(s) + C) ds \\ & + (\phi - \phi_1)^{-1} \cdot (\phi - \phi_1)^{-1} \int_{\phi_1}^{\phi} (\omega_2(s) + C)(s - \phi_1) ds \\ & = M_1 - (\phi - \phi_1)^{-1} M_2 - M_3 + (\phi - \phi_1)^{-1} M_4. \end{aligned}$$

M_1 and M_3 are monotone by virtue of (ϵ) . The integrand in M_2 is the product of two functions of ϕ non-negative and monotone non-decreasing in $[\phi_1, \eta_1]$. Now if two functions $F_1(x), F_2(x)$ are non-negative and monotone non-decreasing in an interval $[a, b]$, then not only are

$$\mathfrak{M}_2(x) = (x - a)^{-1} \int_a^x F_2(s) ds \quad \text{and} \quad \mathfrak{M}_{12}(x) = (x - a)^{-1} \int_a^x F_1(s) F_2(s) ds$$

monotone non-decreasing functions of x there, which is a consequence of (γ) and (ϵ) , but, in addition, as may be proved by using the First Mean Value Theorem, $\mathfrak{M}_{12}(x) = \chi(x) \mathfrak{M}_2(x)$, where $\chi(x)$ is again monotone non-decreasing. If we identify $F_1(x)$ with $\omega_1(\phi) + C$ and $F_2(x)$ with $\phi - \phi_1$, then M_2 corresponds to \mathfrak{M}_{12} and we have

$$M_2(\phi) = \chi(\phi) \mathfrak{M}_2(\phi) = \chi(\phi) (\phi - \phi_1)^{-1} \int_{\phi_1}^{\phi} (s - \phi_1) ds = \frac{1}{2} (\phi - \phi_1) \chi(\phi).$$

⁸ I. e., if $F(x)$ is of bounded variation in $[a, b]$ and $a \leq \xi \leq b$, then $\Phi(\xi) = (\xi - a)^{-1} \int_a^{\xi} F(x) dx$ is of bounded variation in ξ . The proof in the case of a function of bounded variation is readily obtained by decomposing $F(x)$ into its monotone components.

Consequently, $(\phi - \phi_1)^{-1}M_2(\phi) = \frac{1}{2}\chi(\phi)$ and is a monotone (non-decreasing) function of ϕ , and in the same way the monotone character of $(\phi - \phi_1)^{-1}M_4(\phi)$ may be established. Since the sum of a finite number of functions of bounded variation is a function of bounded variation, it follows that the second term in the radicand in the numerator of (14) is of bounded variation in ϕ .

That the second term in the denominator of (14) is of bounded variation in ϕ follows immediately from (ε). From (18) and (β), it follows that the entire numerator of (14) is a function of bounded variation in ϕ . Finally, by virtue of (18), we may apply (α) and (β) and infer that (14), as a function of ϕ , is of bounded variation in $[\phi_1, \eta_1] = [\phi_1(\psi), \eta_1(\psi)]$. Hence, as has been pointed out, $f(t) = f(t; \psi)$ is, for fixed ψ in $[0, 2\pi]$, a function of bounded variation in t in $(0 \leq t \leq a_1(\psi))$.

We now write

$$f(t; \psi) = f(t) = f(+0) + [f(t) - f(+0)].$$

In view of (15), this may be written

$$(21) \quad f(t) = -(-2h_1'')^{-\frac{1}{2}} \\ + [\{(h_1'')^2 + 2h_1''L_{11}\}^{\frac{1}{2}} + h_1'' + L_{10}] / (-2h_1'')^{\frac{1}{2}}(h_1'' + L_{10}) \\ = f(+0) + \Phi(t; \psi).$$

In the first place, we observe that $|f(0)| \geq (2\alpha)^{-\frac{1}{2}}$, uniformly for all ψ . Secondly, $\Phi(t; \psi)$ may be rewritten in the form

$$(22) \quad \Phi(t; \psi) = \frac{2h_1''L_{11} - 2h_1''L_{10} - L_{10}^2}{(-2h_1'')^{\frac{1}{2}}(h_1'' + L_{10})\{[(h_1'')^2 + 2h_1''L_{11}]^{\frac{1}{2}} - h_1'' - L_{10}\}}.$$

From (18) it then follows that the absolute value of the denominator in (22) is not less than $(2\alpha)^{\frac{1}{2}}(\alpha^2/4) = \alpha^{5/2}/2^{3/2} > 0$ uniformly with respect to ϕ , while in the numerator h_1'' is uniformly bounded with respect to ψ and $|L_{jp}| \rightarrow 0$ as $t \rightarrow +0$, uniformly with respect to ψ , as appears from (17). Consequently,

$$(23) \quad |\Phi(t; \psi)| < \epsilon, \text{ if } 0 \leq \phi - \phi_1 < \delta_1(\epsilon); \text{ i. e., if } 0 \leq t < \delta(\epsilon),$$

where δ, δ_1 are independent of ψ .

We now prove, following the method of Hartman, that

If $f(t) = f(t; \psi) = f(+0) + \Phi(t; \psi)$, where $\Phi(t; \psi)$ is of bounded variation in $(0 \leq t \leq a_1(\psi))$ and $\Phi(+0) = 0$, then

$$(24) \quad \int_0^{a_1(\psi)} f(t; \psi) \exp[-irt^2] dt = \frac{1}{2}f(+0)\Gamma(\frac{1}{2})(ir)^{-\frac{1}{2}} + o(r^{-\frac{1}{2}}),$$

where the o -term holds uniformly for all ψ .

Proof:

$$(25) \quad \int_0^{a_1(\psi)} f(t; \psi) \exp[-irt^2] dt = \int_0^{a_1(\psi)} f(+0) \exp[-irt^2] dt \\ + \int_0^{a_1(\psi)} \Phi(t; \psi) \exp[-irt^2] dt = Ia + Ib.$$

Now ⁴

$$Ia = f(+0) \int_0^{+\infty} \exp[-irt^2] dt - f(+0) \int_{a_1(\psi)}^{+\infty} \exp[-irt^2] dt \\ = \frac{1}{2} f(+0) r^{-\frac{1}{2}} \pi^{\frac{1}{2}} \exp[-i\pi/4] - f(+0) \int_{a_1(\psi)}^{+\infty} \exp[-irt^2] dt.$$

But $|f(+0)| = (-2h_1'')^{-\frac{1}{2}} \leq (2\alpha)^{-\frac{1}{2}}$ for all ψ . Furthermore,

$$(26) \quad \left| \int_{a_1(\psi)}^{+\infty} \exp[-irt^2] dt \right| < \text{Const.}/r.$$

For $G_2(r) = \int_r^{+\infty} y^{-\frac{1}{2}} \exp[-iy] dy$ exists and is $O(r^{-\frac{1}{2}})$ in virtue of the Second Mean Value Theorem applied to a finite interval. On setting $rt^2 = y$, the integral in (26) becomes, up to a constant factor,

$$r^{-\frac{1}{2}} G_2(r[a_1(\psi)]^2) = O(r^{-1}), \text{ since } a_1(\psi) > \beta > 0 \text{ for all } \psi,$$

where β is the constant defined above following equation (6). Consequently, by (21) and (26),

$$(27) \quad Ia = -\frac{1}{2} (-2h_1''r)^{-\frac{1}{2}} (\pi)^{\frac{1}{2}} \exp[-i\pi/4] + O(r^{-1}),$$

where the O -term is independent of ψ , in the sense that in absolute value it is not greater than $\text{const.}/r$, where the constant is *independent of* ψ .

It therefore remains to consider Ib , where $\Phi(t; \psi)$ is of bounded variation in $(0 \leq t \leq a_1(\psi))$ and $\Phi(+0; \psi) = 0$ uniformly with respect to ψ in the sense that

$$|\Phi(t; \psi)| < \epsilon \text{ for all } t, 0 \leq t < \delta, \delta = \delta(\epsilon) \text{ independent of } \psi.$$

We next define the non-increasing function $m(r)$ by

$$(28) \quad m(r) = \text{l. u. b. } |\Phi(t; \psi)| \text{ for } 0 < t \leq r^{-1}; 0 \leq \psi \leq 2\pi,$$

so that $m(r) \rightarrow 0$ as $r \rightarrow +\infty$. Since we are interested only in very large values of r , we may always suppose $0 < r^{-1} \leq \beta < a_1(\psi)$ for all ψ . Let $\lambda(r)$ be a non-decreasing function of r which becomes infinite with r so slowly that

$$(29) \quad m[r^{\frac{1}{2}}(\lambda(r))^{-1}] \lambda(r) \rightarrow 0, \quad (r \rightarrow +\infty).$$

In particular, we may let $\lambda(r) = \min(r^{1/4}, (m[r^{1/4}])^{-1})$. Now

$$(30) \quad \int_0^{a_1(\psi)} \Phi(t; \psi) \exp[-irt^2] dt = \frac{1}{2} \int_0^{a_1^2(\psi)} \Phi(t^{\frac{1}{2}}; \psi) t^{\frac{1}{2}} \exp[-irt] dt.$$

Consider the last integral from 0 to b , where $0 < b \leq a_1^2(\psi)$.

$$(31) \quad \left| \int_0^b \Phi(t^{\frac{1}{2}}; \psi) t^{\frac{1}{2}} \exp[-irt] dt \right| \leq m(b^{-\frac{1}{2}}) \int_0^b t^{\frac{1}{2}} dt = m(b^{-\frac{1}{2}}) \cdot 2b^{\frac{3}{2}}.$$

If we place $b = r^{-1}(\lambda(r))^2$, the last expression on the right of (31) becomes

$$2m[r^{\frac{1}{2}}/\lambda(r)]\lambda(r)r^{\frac{3}{2}} = o(r^{-\frac{1}{2}})$$

by virtue of (29). Hence

$$(32) \quad \int_0^b \Phi(t^{\frac{1}{2}}; \psi) t^{\frac{1}{2}} \exp[-irt] dt = o(r^{-\frac{1}{2}}) = \zeta(r)r^{-\frac{1}{2}},$$

where $|\zeta(r)| < \epsilon$ if $r \geq R(\epsilon)$, R independent of ψ , since $m(\cdot)$ is by definition independent of ψ and λ depends only on r and on $m(\cdot)$.

In order to appraise $\int_b^{a_1^2(\psi)} \Phi(t^{\frac{1}{2}}; \psi) t^{\frac{1}{2}} \exp[-irt] dt$, we apply the Second Mean Value Theorem to the monotone function $t^{\frac{1}{2}}$, obtaining

$$(33) \quad b^{-\frac{1}{2}} \int_b^{\xi} \Phi(t^{\frac{1}{2}}; \psi) \exp[-irt] dt + [a_1(\psi)]^{-1} \int_{\xi}^{a_1^2(\psi)} \Phi(t^{\frac{1}{2}}; \psi) \exp[-irt] dt,$$

where it is understood that the Second Mean Value Theorem is applied separately to the real and the imaginary parts of the integral, the notation being

$$\int_{\xi}^{\eta} (\cdots) dx = \int_{\xi_1}^{\xi_2} R(\cdots) dx + i \int_{\xi_2}^{\xi_1} I(\cdots) dx, \quad (b < \xi_1 < a_1^2(\psi); b < \xi_2 < a_1^2(\psi)).$$

Now $\Phi(t^{\frac{1}{2}}; \psi)$ is of bounded variation, inasmuch as $\Phi(t; \psi)$ is, so it may be supposed without loss of generality that $\Phi(t^{\frac{1}{2}}; \psi)$ is a bounded monotone function, whereupon the Second Mean Value may be applied to each of the integrals in (33). From (17) and the continuity of $h''(\phi; \psi)$, hence of $\omega(\phi; \psi)$, on the torus and consequently for $0 \leq t \leq a_1(\psi)$ or $0 \leq \tau \leq a_1^2(\psi)$, where τ is the t of the right-hand member of (30), and for $0 \leq \psi < 2\pi$, it follows that L_{10} and L_{11} are bounded in $0 \leq \tau \leq a_1^2(\psi)$ uniformly with respect to ψ . Therefore, from (22) and the remark immediately following it, one infers the existence of a constant K such that

$$|\Phi(t^{\frac{1}{2}}; \psi)| < K \text{ for all } t, 0 \leq t \leq a_1^2(\psi), \text{ and all } \psi \text{ in } (0 \leq \psi < 2\pi).$$

Finally, $0 < b < a_1^2(\psi)$, so that $[a_1(\psi)]^{-1} \leq b^{-\frac{1}{2}}$ and from (33) it follows that

$$(34) \quad \left| \int_b^{a_1^2(\psi)} \Phi(t^{\frac{1}{2}}; \psi) t^{\frac{1}{2}} \exp[-irt] dt \right| \leq 16Kb^{-\frac{1}{2}}r^{-1} = 16K[\lambda(r)]^{-1}r^{-\frac{1}{2}} = o(r^{-\frac{1}{2}}),$$

where the o -term is uniform with respect to ψ in the same sense as in (32). From (25), (27), (30), (32) and (34), it then follows that

$$\int_0^{a_1(\psi)} f(t; \psi) \exp[-irt^2] dt = -\frac{1}{2}(-2h_1''r)^{-\frac{1}{2}}(\pi)^{\frac{1}{2}} \exp[-i\pi/4] + o(r^{-\frac{1}{2}}),$$

corresponding to (24).

Substituting into (8), we obtain

$$(35) \quad J_1 = \frac{1}{2}(2\pi r)^{-\frac{1}{2}}[-h''(\phi_1(\psi); \psi)]^{-\frac{1}{2}} \exp[i(rh(\phi_1(\psi); \psi) - \pi/4)] + o(r^{-\frac{1}{2}}),$$

the o -term being uniform with respect to ψ .

To calculate the integral J_2 , we observe that $h'(\phi; \psi)$ is negative for $\eta_1(\psi) \leq \phi \leq \eta_2(\psi)$, so that $h(\eta_1(\psi); \psi) - h(\phi; \psi)$ is in this interval steadily increasing from zero, and if we set

$$t = |h(\eta_1(\psi); \psi) - h(\phi; \psi)|^{\frac{1}{2}},$$

t increases from 0 to $a_2(\psi) = |h(\eta_1(\psi); \psi) - h(\eta_2(\psi); \psi)|^{\frac{1}{2}}$ as ϕ increases from $\eta_1(\psi)$ to $\eta_2(\psi)$. By the introduction of t as integration variable in J_2 ,

$$J_2 = -\frac{1}{\pi} \exp[irh(\eta_1(\psi); \psi)] \int_0^{a_2(\psi)} \exp[-irt^2] t/h'(\phi(t, \psi); \psi) dt.$$

This last integral is of the form $\int_0^{a_2(\psi)} f(t; \psi) \exp[-irt^2] dt$, where

$$(36) \quad f = f(t; \psi) = t/h'(\phi(t, \psi); \psi).$$

Just as $\eta_1(\psi)$ has already been so chosen that $\xi/3 < \eta_1(\psi) - \phi_1(\psi) < 2\xi/3$, one may so choose $\eta_2(\psi)$ that $\xi/3 < \phi_3(\psi) - \eta_2(\psi) < 2\xi/3$, where ξ (defined just above equation (4)) is independent of ψ . Then from continuity considerations it follows that $h'(\phi(t, \psi); \psi) > \gamma_1 > 0$ for all t in $(0 \leq t \leq a_2(\psi) \leq (2\mu)^{\frac{1}{2}})$ and for all ψ , μ being the maximum of $h(\phi; \psi)$ on the torus. Since $h''(\phi; \psi)$ is continuous and of bounded variation in ϕ , $h'(\phi; \psi)$ enjoys the same properties. Moreover, ϕ is a continuous monotone non-decreasing function of t , so that $h'(\phi(t, \psi); \psi)$, as a function of t in $(0 \leq t \leq a_2(\psi))$ is of bounded variation in t . Consequently, by (β) and (α) , $f(t)$ is, for fixed ψ , a function of bounded variation in t . Moreover, $f(0; \psi) = 0/h'(\eta_1(\psi); \psi) = 0$ for all ψ , so that, in J_2 , $f(t; \psi)$ plays the rôle of $f(t; \psi) - f(0; \psi) = \Phi(t; \psi)$ in J_1 , and, inasmuch as

$$(37) \quad |f(t; \psi)| \leq t/\gamma_1 \leq (2\mu)^{\frac{1}{2}}/\gamma_1,$$

it follows on the one hand that $|f(t; \psi)| < \epsilon$ for all t such that $0 \leq t < \delta_2 = \delta_2(\epsilon)$, where $\delta_2(\epsilon)$ is independent of ψ , while, on the other hand,

there exists a constant K_1 such that $|f(t; \psi)| < K_1$ for all t in $(0 \leq t \leq a_2(\psi))$ and all ψ . By the same reasoning as that used in the calculation of J_1 , it then follows that

$$(38) \quad J_2 = o(r^{-\frac{1}{2}}),$$

where the o -term is independent of ψ .

To each zero of h' of the form ϕ_{4k-3} , ($k = 1, \dots, n/2$), there correspond two integrals, of which one, like J_1 , has ϕ_{4k-3} as lower integration limit, while the other, like J_{3n} , has $\phi_{4k-3} (= \phi_{4k-3} + 2\pi)$ as upper integration limit. The contribution of order $r^{-\frac{1}{2}}$ from each of these may readily be shown to be the same, viz.

$$(39) \quad \frac{1}{2}(2\pi r)^{-\frac{1}{2}}[-h''(\phi_{4k-3}(\psi); \psi)]^{-\frac{1}{2}} \exp[i(rh(\phi_{4k-3}(\psi); \psi) - \pi/4)].$$

Similarly, by a slight modification of the foregoing reasoning, it may be proved that to each zero of h' of the form ϕ_{4k-1} , ($k = 1, 2, \dots, n/2$), there correspond two integrals, from each of which the contribution of order $r^{-\frac{1}{2}}$ is

$$(40) \quad \frac{1}{2}(2\pi r)^{-\frac{1}{2}}[h''(\phi_{4k-1}(\psi); \psi)]^{-\frac{1}{2}} \exp[i(rh(\phi_{4k-1}(\psi); \psi) + \pi/4)].$$

Finally, just as in the case of J_2 , it may be shown that for each of the integrals J_{3k-1} , ($k = 1, \dots, n$), over an interval containing a zero of h'' ,

$$(41) \quad J_{3k-1} = o(r^{-\frac{1}{2}}),$$

where the o -term is independent of ψ .

From (39), (40) and (41), we then obtain (3), q. e. d.

THE LINCOLN UNIVERSITY,
CHESTER COUNTY, PENNSYLVANIA.

ON TAUBERIAN THEOREMS FOR DOUBLE SERIES.*¹

By RALPH PALMER AGNEW.

1. Introduction. Let

$$s_n = \sum_{k=1}^n u_k; \quad \sigma_n = \frac{1}{n} \sum_{k=1}^n s_k \quad (n = 1, 2, \dots)$$

denote the sequence of partial sums and the C_1 transform of a real series Σu_n . A classic Tauberian theorem states that if $\sigma_n \rightarrow s$ and the unilateral Tauberian condition $nu_n < K$ is satisfied, then $s_n \rightarrow s$.

Let

$$s_{m,n} = \sum_{j,k=1}^{m,n} u_{j,k}; \quad \sigma_{m,n} = \frac{1}{mn} \sum_{j,k=1}^{m,n} s_{j,k} \quad (m, n = 1, 2, \dots)$$

denote the sequence of partial sums and the C_1 transform of a real double series $\Sigma u_{m,n}$. K. Knopp² has recently proved several Tauberian theorems of which his third is the following:

If $\sigma_{m,n} \rightarrow s$ and $(m^2 + n^2)u_{m,n} < K$, then $s_{m,n} \rightarrow s$.

The "natural" question whether this theorem holds when the Tauberian condition $(m^2 + n^2)u_{m,n} < K$ is replaced by the weaker condition $mn u_{m,n} < K$ was raised and left unanswered by Knopp.

In § 2 we give examples which show that the unilateral condition $mn u_{m,n} < K$ will not serve; that the stronger O -condition $mn | u_{m,n} | < K$ will not serve; and in fact that the still stronger set of o -conditions

$$(1) \quad \begin{aligned} \lim_{n \rightarrow \infty} mn | u_{m,n} | &= 0 & (m = 1, 2, \dots) \\ \lim_{m \rightarrow \infty} mn | u_{m,n} | &= 0 & (n = 1, 2, \dots) \\ \lim_{m, n \rightarrow \infty} mn | u_{m,n} | &= 0 \end{aligned}$$

will not serve. The sequences d_n and ϵ_n of § 2 are specialized to obtain further results of this character.

In § 3, we show that the situation is the same for many other methods

* Received December 7, 1939.

¹ Presented to the American Mathematical Society, February 24, 1940.

² K. Knopp, "Limitierungs-Umkehrsätze für Doppelfolgen," *Mathematische Zeitschrift*, vol. 45 (1939), pp. 573-589, p. 581. Adjustment from Knopp's subscripts 0, 1, 2, . . . to our subscripts 1, 2, 3, . . . is easily made.

of summability, including the Cesàro methods of all positive orders and the Abel power series method.

In § 4, we show that the stronger hypothesis that all of the limits

$$\lim_{n \rightarrow \infty} \sigma_{mn}; \quad \lim_{m \rightarrow \infty} \sigma_{mn}; \quad \lim_{m, n \rightarrow \infty} \sigma_{mn}$$

exist, the first for each $m = 1, 2, \dots$ and the second for each $n = 1, 2, \dots$, together with a Tauberian condition such as (1), implies neither convergence nor convergence by rows of Σu_{mn} .

It therefore appears that the double sequence $mn u_{m,n}$ does not play, in Tauberian theory for double series, a rôle analogous to the rôle of the simple sequence nu_n in Tauberian theory for simple series.

In connection with the examples of § 2, it is illuminating (but not essential) to recognize the fact that if $\Sigma u_{m,n}$ has bounded partial sums $s_{m,n}$ and $s_{m,n} \rightarrow s$, then $\sigma_{m,n} \rightarrow s$; and that, irrespective of whether $\Sigma u_{m,n}$ has bounded partial sums, if $s_{m,n} \rightarrow s$ and $\sigma_{m,n} \rightarrow \sigma$, then $s = \sigma$. General theory and references to literature covering these points may be found in two papers in this Journal.³ It follows that if $\sigma_{m,n} \rightarrow s$ and it is not true that $s_{m,n} \rightarrow s$, then $\lim s_{m,n}$ cannot exist.

2. Some examples. Let d_n be a bounded sequence of real non-negative numbers such that $\Sigma d_n = \infty$. Let ϵ_n be a sequence of positive numbers for which $0 < \epsilon_n \leq 1$. Choose D such that

$$0 \leq d_n < D \quad (n = 1, 2, 3, \dots),$$

and let $n_0 = 0$.

We define by induction a sequence

$$n_0 < v_1 < n_1 < v_2 < n_2 < v_3 < n_3 < \dots$$

of indices and the terms u_{mn} of a series Σu_{mn} . For the first step in the induction, take $k = 1$. Define $u_{2k-1,n}$ for $n = 1, 2, 3, \dots$ by the formulas

$$(2) \quad \begin{aligned} u_{2k-1,n} &= \epsilon'_k d_n & n_{k-1} < n \leq v_k \\ &= -\epsilon'_k d_n & v_k < n < n_k \\ &= -\epsilon'_k \theta_k d_n & n = n_k \\ &= 0 & \text{otherwise} \end{aligned}$$

where ϵ'_k is the lesser of ϵ_{2k-1} and ϵ_{2k} ; v_k is so chosen that

$$(3) \quad D < \epsilon'_k (d_{n_{k-1}+1} + d_{n_{k-1}+2} + \dots + d_{v_k}) < 2D;$$

³ R. P. Agnew, "On summability of double sequences," *American Journal of Mathematics*, vol. 54 (1932), pp. 648-656; "On summability of multiple sequences," *ibid.*, vol. 56 (1934), pp. 62-68.

n_k is so chosen that

$$(4) \quad \epsilon'_k(d_{n_{k-1}+1} + \cdots + d_{v_k}) - \epsilon'_k(d_{v_{k+1}} + \cdots + d_{q-1} + \theta d_q)$$

is ≥ 0 when $\theta = 1$ and $q = n_k - 1$ but is < 0 when $\theta = 1$ and $q = n_k$; and finally θ_k is chosen such that the difference (4) is 0 when $q = n_k$ and $\theta = \theta_k$. Observe that $0 \leq \theta_k < 1$. Let $u_{2k,n}$ be defined for $n = 1, 2, 3, \cdots$ by the formulas

$$(5) \quad u_{2k,n} = -u_{2k-1,n} \quad (n = 1, 2, 3, \cdots).$$

Successive steps in the induction are obtained by giving k the values 2, 3, 4, \cdots in turn.

The terms of the series $\Sigma u_{m,n}$ which we have just defined may be displayed in the form

$$(6) \quad \begin{array}{cccccccc} x & + & \cdots & + & x & + & 0 & + & \cdots & + & 0 & + & 0 & + & \cdots & + & 0 & + & \cdots \\ y & + & \cdots & + & y & + & 0 & + & \cdots & + & 0 & + & 0 & + & \cdots & + & 0 & + & \cdots \\ 0 & + & \cdots & + & 0 & + & x & + & \cdots & + & x & + & 0 & + & \cdots & + & 0 & + & \cdots \\ 0 & + & \cdots & + & 0 & + & y & + & \cdots & + & y & + & 0 & + & \cdots & + & 0 & + & \cdots \\ 0 & + & \cdots & + & 0 & + & 0 & + & \cdots & + & 0 & + & x & + & \cdots & + & x & + & \cdots \\ 0 & + & \cdots & + & 0 & + & 0 & + & \cdots & + & 0 & + & y & + & \cdots & + & y & + & \cdots \\ & & & & & & & & & & & & & & & & & & + & \cdots \end{array}$$

in which the value of each $u_{m,n}$ which may differ from 0 is represented by an x or by a y . The definition of θ_k implies that the sum of the x 's in each row is 0, and (5) implies that each y is the negative of the x above it. These considerations imply that the sequence $s_{m,n}$ of partial sums of the series $\Sigma u_{m,n}$ may be displayed in the form

$$(7) \quad \begin{array}{cccccccc} z, & \cdots, & z, & 0, & \cdots, & 0, & 0, & \cdots, & 0, & \cdots \\ 0, & \cdots, & 0, & 0, & \cdots, & 0, & 0, & \cdots, & 0, & \cdots \\ 0, & \cdots, & 0, & z, & \cdots, & z, & 0, & \cdots, & 0, & \cdots \\ 0, & \cdots, & 0, & 0, & \cdots, & 0, & 0, & \cdots, & 0, & \cdots \\ 0, & \cdots, & 0, & 0, & \cdots, & 0, & z, & \cdots, & z, & \cdots \\ 0, & \cdots, & 0, & 0, & \cdots, & 0, & 0, & \cdots, & 0, & \cdots \\ \cdots & & & & & & & & & \end{array}$$

in which the value of each $s_{m,n}$ which may differ from 0 is represented by a z .

The definitions of v_k , n_k , and $u_{m,n}$ imply that

$$(8) \quad 0 \leq s_{m,n} < 2D \quad (m, n = 1, 2, \cdots),$$

$$(9) \quad D < s_{2k-1, v_k} < 2D \quad (k = 1, 2, \cdots),$$

$$(10) \quad s_{2k,n} = 0 \quad (k, n = 1, 2, \dots).$$

Hence

$$(11) \quad \liminf_{m, n \rightarrow \infty} s_{m,n} = 0; \quad D \leq \limsup_{m, n \rightarrow \infty} s_{m,n} \leq 2D$$

and therefore $\lim s_{m,n}$ does not exist.

The fact that $0 \leq s_{m,n} \leq 2D$, and that at most n of the terms $s_{j,k}$ in the sum

$$\sigma_{m,n} = \frac{1}{mn} \sum_{j,k=1}^{m,n} s_{j,k}$$

are different from 0, implies that $0 \leq \sigma_{m,n} \leq 2D/m$ and hence that $\sigma_{m,n} \rightarrow 0$.

Our definitions imply that, for each k and n ,

$$|u_{2k,n}| = |u_{2k-1,n}| \leq \epsilon'_k d_n;$$

and since ϵ'_k is the lesser of ϵ_{2k-1} and ϵ_{2k} , this implies that

$$(12) \quad |u_{m,n}| \leq \epsilon_m d_n \quad (m, n = 1, 2, \dots).$$

For the particular sequences

$$(13) \quad d_n = 1/n \log(n+1); \quad \epsilon_n = 1/n 2^n$$

the series Σu_{mn} satisfies the Tauberian condition

$$(14) \quad mn |u_{m,n}| \leq 1/2^m \log(n+1)$$

while $\sigma_{m,n} \rightarrow 0$ and the sequence $s_{m,n}$ is bounded and $\lim s_{m,n}$ fails to exist.

For the sequences

$$(15) \quad d_n = \epsilon_n = 1/n [\log(n+2)] [\log \log(n+16)]$$

we obtain the symmetric inadequate Tauberian condition

$$(16) \quad mn \log(m+2) \log(n+2) |u_{m,n}| \leq 1/\log \log(m+16) \log \log(n+16).$$

Each one of (14) and (16) demonstrates inadequacy of the σ -conditions (1).⁴

3. Other methods of summability. Let $a_{n,k}$ and $b_{n,k}$, $n, k = 1, 2, 3, \dots$, denote matrices of regular simple-sequence transformations

$$(17) \quad \sigma_n^{(a)} = \sum_{k=1}^{\infty} a_{n,k} s_k; \quad \sigma_n^{(b)} = \sum_{k=1}^{\infty} b_{n,k} s_k;$$

and let the matrix $a_{n,k}$ satisfy the additional condition

⁴ An example of a divergent series Σu_{mn} which is summable C_1 , and which satisfies the condition $mn |u_{mn}| < K$ and the condition $mn u_{mn} \rightarrow 0$ as $m, n \rightarrow \infty$, has just been published by W. Meyer-König, "Zur Frage der Umkehrung des C' - und A -Verfahrens bei Doppelfolgen," *Mathematische Zeitschrift*, vol. 46 (1940), pp. 157-160.—Added in the proof.

$$(18) \quad \lim_{n \rightarrow \infty} \text{l. u. b. } |a_{n,k}| = 0.$$

This condition is of course not satisfied when $a_{n,k}$ is the identity matrix $\delta_{n,k}$, but it is satisfied for many other regular matrices. In particular, (18) is satisfied when $a_{n,k}$ is the matrix

$$a_{n,k}^{(r)} = \begin{cases} \frac{r}{n} \left(\frac{n}{n+r-1} \right) \left(\frac{n-1}{n+r-2} \right) \cdots \left(\frac{n-k+1}{n+r-k} \right) & 1 \leq k \leq n \\ = 0 & k > n \end{cases}$$

of a Cesàro transformation C_r whose order r is a real or complex number with a positive real part r' ; for

$$|a_{n,k}^{(r)}| \leq \frac{|r|}{n} \left(\frac{n}{n+r'-1} \right) \left(\frac{n-1}{n+r'-2} \right) \cdots \left(\frac{n-k+1}{n+r'-k} \right)$$

when $1 \leq k \leq n$ so that if $r' \geq 1$,

$$|a_{n,k}^{(r)}| \leq |r|/n \quad (k = 1, 2, \dots),$$

and if $0 < r' < 1$

$$|a_{n,k}^{(r)}| \leq |r| \Gamma(r') \Gamma(n) / \Gamma(n+r') \quad (k = 1, 2, \dots).$$

It is well known that C_r is regular when $r' > 0$.

Let $A \odot B$ denote the double sequence method of summability defined by

$$(19) \quad \sigma_{m,n}^{(a,b)} = \sum_{j,k=1}^{\infty} a_{m,j} b_{n,k} s_{j,k}.$$

Let $\Sigma u_{m,n}$ and $s_{m,n}$ be as constructed in § 2. Then $|s_{m,n}| < 2D$ so that the series in (19) converges absolutely; hence

$$\sigma_{m,n}^{(a,b)} = \sum_{k=1}^{\infty} b_{n,k} \sum_{j=1}^{\infty} a_{m,j} s_{j,k}.$$

For each k there is at most one j , say β_k , for which $s_{j,k} \neq 0$. Hence

$$(21) \quad \sigma_{m,n}^{(a,b)} = \sum_{k=1}^{\infty} b_{n,k} a_{m,\beta_k} s_{\beta_k,k}$$

so that

$$(22) \quad |\sigma_{m,n}^{(a,b)}| \leq \sum_{k=1}^{\infty} |b_{n,k}| [\text{l. u. b. } |a_{m,i}|] [2D]$$

and therefore

$$(23) \quad \lim_{m,n \rightarrow \infty} \sigma_{m,n}^{(a,b)} = 0.$$

Thus $\Sigma u_{m,n}$ is summable $A \odot B$ to 0, and the examples of § 2 apply to $A \odot B$ as well as to C_1 . In particular the examples apply to the Cesàro transformation

$C_r \odot C_s$ if the real parts of r and s are positive. In case $r=s=1$, $C_r \odot C_s$ becomes the special double sequence transformation C_1 previously considered.

It can be shown in the same way that if

$$(24) \quad \sigma^{(a)}(t) = \sum_{k=1}^{\infty} a_k(t) s_k; \quad \sigma^{(b)}(t) = \sum_{k=1}^{\infty} b_k(t) s_k$$

are regular sequence-to-function transformations and

$$(25) \quad \lim_{t \rightarrow t_0} \text{l. u. b. } |a_k(t)| = 0, \quad k=1, 2, \dots$$

then each series $\Sigma u_{m,n}$ of § 2 is summable to 0 by the double sequence-to-function transformation

$$(26) \quad \sigma^{(a,b)}(t, u) = \sum_{j,k=1}^{\infty} a_j(t) b_k(u) s_{j,k}.$$

This applies to the Abel power series method for which

$$(27) \quad a_k(t) = b_k(t) = t^{k-1}(1-t),$$

the variable t approaching 1 over the real interval $0 \leq t < 1$ or over the complex sets of Stolz and Pringsheim.

4. Convergence by rows. A double series is called *convergent by rows* to s_R if

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} u_{m,n} = s_R$$

or, what amounts to the same thing, if $\lim_{n \rightarrow \infty} s_{m,n}$ exists for each $m=1, 2, \dots$ and

$$(28) \quad \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} s_{m,n} = s_R.$$

The series constructed in § 2 converge by rows to 0; hence the examples do not preclude the possibility that $\sigma_{m,n} \rightarrow s$ and the Tauberian condition $mn u_{m,n} < K$ may imply (28) or at least the weaker condition

$$(29) \quad \lim_{m \rightarrow \infty} \liminf_{n \rightarrow \infty} s_{m,n} = \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} s_{m,n} = s.$$

This question and others are settled by the following example.

Let d_n , ϵ_n , and D be given as in § 2; and for each $k=1, 2, \dots$ let ϵ'_k be the least of the four numbers ϵ_{4k-3} , ϵ_{4k-2} , ϵ_{4k-1} , and ϵ_{4k} . For each $k=1, 2, \dots$, choose n_k such that

$$D < \epsilon'_k(d_1 + d_2 + \dots + d_{n_k}) < 2D$$

and let

$$u_{4k-3,n} = -u_{4k-2,n} = -u_{4k-1,n} = u_{4k,n} \quad (n=1, 2, \dots)$$

where

$$\begin{aligned} u_{4k,n} &= \epsilon'_k d_n & 1 \leq n \leq n_k, \\ &= 0 & n > n_k. \end{aligned}$$

The sequence $s_{m,n}$ of partial sums, and the transforms by various methods of summability, of this series are more complicated than those for the series of § 2. However it is possible to show that $\Sigma u_{m,n}$ satisfies the Tauberian condition

$$|u_{m,n}| \leq \epsilon_m d_n;$$

that $-2D \leq s_{m,n} \leq 2D$; that if $\sigma_{m,n}$ is as before the C_1 transform of $\Sigma u_{m,n}$, then

$$(30) \quad \lim_{n \rightarrow \infty} \sigma_{m,n}, \quad \lim_{m \rightarrow \infty} \sigma_{m,n}, \quad \lim_{m, n \rightarrow \infty} \sigma_{m,n}$$

all exist, the first for each $m = 1, 2, \dots$ and the second for each $n = 1, 2, \dots$; that $\Sigma u_{m,n}$ fails to converge; and finally that each row of the series $\Sigma u_{m,n}$ converges but that the series of values of the rows does not converge.

This example is of interest because existence of the first limits in (30) and the Tauberian condition $mnu_{m,n} < K$ imply (by iterated use of the Tauberian theorem for simple series given in § 1) convergence of each row of $\Sigma u_{m,n}$. The example shows that existence of all of the limits in (30) and stronger Tauberian conditions $|u_{m,n}| \leq \epsilon_m d_n$ do not imply convergence of the series of values of the rows.

5. Conclusion. It is sometimes desirable to have, in addition to a proof of a result, a plausible argument which indicates roughly why the result may possibly hold. The question here is "why" $\sigma_{m,n} \rightarrow s$ and $mnu_{m,n} < K$ can fail to imply $s_{m,n} \rightarrow s$ as $\sigma_n \rightarrow s$ and $nu_n < K$ imply $s_n \rightarrow s$. The "answer" seems to be that the condition $mnu_{m,n} < K$ does not prevent an effective dilution of a double sequence $s_{m,n}$ by insertion of zeros in the two dimensional pattern, while $nu_n < K$ does prevent an effective dilution of a simple sequence s_n by insertion of zeros in the linear pattern.

CORNELL UNIVERSITY,
ITHACA, NEW YORK.

ANALYTIC FUNCTIONS AND MULTIPLE FOURIER INTEGRALS.*

By W. T. MARTIN.

Introduction. In the first part of this note we consider the class **E** of entire functions $f(z_1, \dots, z_n)$ which satisfies relations of the form

$$(1) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |f(x_1 + iy_1, \dots, x_n + iy_n)|^2 dx_1 \cdots dx_n < A e^{2a(|y_1| + \dots + |y_n|)}$$

for all finite values of y_1, \dots, y_n where a and A are positive constants. It is easily shown that this class of functions is identical with the class of functions having Fourier transforms $\phi(u_1, \dots, u_n) e^{(y_1 u_1 + \dots + y_n u_n)}$ which vanish outside a certain finite region. Next if we denote by K the common part of all convex bodies (in the u -space) in whose exteriors ϕ vanishes identically and by $s(\lambda)$ its supporting function,

$$(2) \quad s(\lambda) = \max_{(u) \in K} \{\lambda_1 u_1 + \dots + \lambda_n u_n\}, \quad \lambda_1, \dots, \lambda_n \text{ real},$$

then we show that $s(\lambda)$ is equal to a growth-function $h(\lambda)$ of f defined as follows¹

$$(3) \quad h(\lambda) = \frac{1}{2} \lim_{\rho \rightarrow \infty} \frac{1}{\rho} \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |f(x_1 + i\lambda_1 \rho, \dots, x_n + i\lambda_n \rho)|^2 dx_1 \cdots dx_n.$$

From these considerations it follows that the class **E** is identical with the class considered by Plancherel and Pólya² of entire functions of integrable square over the real space $y_1 = \dots = y_n = 0$ and that the growth-function $h(\lambda)$ defined in (3) is equal to the growth-function

$$(4) \quad h_P(\lambda) = \max_{\substack{-\infty < a_k < \infty \\ k=1, \dots, n}} \limsup_{\rho \rightarrow \infty} \frac{1}{\rho} \log |f(\alpha_1 + i\lambda_1 \rho, \dots, \alpha_n + i\lambda_n \rho)|$$

defined by them.

In the second section we prove results of a similar nature for the class of functions f analytic in the "octant" $I_m\{z_k\} > 0, k = 1, \dots, n$, and satisfying relations of the form (1) for all positive values of y_1, \dots, y_n .

* Received October 12, 1939.

¹ The idea of considering a growth-function of the sort defined here arose in a conversation which the author had with Professor S. Bochner.

² M. Plancherel and G. Pólya, "Fonctions entières et integrales de Fourier multiples," *Commentarii Math. Helvetici*, vol. 9 (1936-37), pp. 224-248; vol. 10 (1937-38), pp. 110-163.

1. **The class P of entire functions.** We consider the class P of functions f representable in the form

$$(5) \quad f(z_1, \dots, z_n) = \left(\frac{1}{2\pi}\right)^{n/2} \int_{-a}^a \dots \int_{-a}^a \phi(u) e^{-iu_1 z_1 - \dots - iu_n z_n} d\omega_u,$$

where $\phi(u) \equiv \phi(u_1, \dots, u_n)$ is of integrable square over $-a < u_k < a$, $k = 1, \dots, n$, and $d\omega_u$ is the volume element $du_1 \dots du_n$, and we show that this class is identical with the class E of entire functions satisfying relations of the form (1). First, by the Schwarz inequality,

$$\begin{aligned} & \left| \int_{-a}^a \dots \int_{-a}^a \phi(u) e^{-iu_1 z_1 - \dots - iu_n z_n} d\omega_u \right|^2 \\ & \leq \int_{-a}^a \dots \int_{-a}^a |\phi(u)|^2 d\omega_u \int_{-a}^a \dots \int_{-a}^a e^{2(y_1 u_1 + \dots + y_n u_n)} d\omega_u \end{aligned}$$

and thus the function f defined in (5) is an entire function. Next by Plancherel's theorem

$$\begin{aligned} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1 + iy_1, \dots, x_n + iy_n)|^2 d\omega_x &= \int_{-a}^a \dots \int_{-a}^a |\phi(u)|^2 e^{2(y_1 u_1 + \dots + y_n u_n)} d\omega_u \\ &\leq e^{2a(|y_1| + \dots + |y_n|)} \int_{-a}^a \dots \int_{-a}^a |\phi|^2 d\omega_u, \end{aligned}$$

and thus a relation of the form (1) holds. Conversely, if f belongs to the class E , then for each (y_1, \dots, y_n) it has a Fourier transform $\psi_{(y)}(u)$. By a theorem due to Bochner,³ since the left-hand side of (1) is bounded for (y) in any bounded region, it follows that $\psi_{(y)}(u)$ has the form $\phi(u_1, \dots, u_n) e^{y_1 u_1 + \dots + y_n u_n}$. Thus by Plancherel's theorem

$$(6) \quad f(z_1, \dots, z_n) \sim \left(\frac{1}{2\pi}\right)^{n/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \phi(u) e^{y_1 u_1 + \dots + y_n u_n} e^{-iu_1 z_1 - \dots - iu_n z_n} d\omega_u.$$

We next show that if f has the representation (6) and if (1) holds then $\phi \equiv 0$ outside the "cube" C $[-a < u_k < a, k = 1, \dots, n]$. For suppose $\phi \not\equiv 0$ in some region R which lies outside C . There is no loss in generality in assuming that R is of the form $\alpha_k < u_k < \beta_k, k = 1, \dots, n$, where $a < \alpha_1 < \beta_1$. Then for y_1, \dots, y_n positive Plancherel's theorem yields

$$\begin{aligned} (7) \quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1 + iy_1, \dots, x_n + iy_n)|^2 d\omega_x &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |\phi|^2 e^{2(y_1 u_1 + \dots + y_n u_n)} d\omega_u \\ &\geq \int_{\alpha_1}^{\beta_1} \dots \int_{\alpha_n}^{\beta_n} |\phi|^2 e^{2(y_1 u_1 + \dots + y_n u_n)} d\omega_u \\ &\geq e^{2(\alpha_1 y_1 + \dots + \alpha_n y_n)} \int_{\alpha_1}^{\beta_1} \dots \int_{\alpha_n}^{\beta_n} |\phi|^2 d\omega_u. \end{aligned}$$

³ S. Bochner, "Bounded analytic functions in several variables and multiple Laplace integrals," *American Journal of Mathematics*, vol. 59 (1937), pp. 732-738, esp. 733-734.

As $y_1 \rightarrow \infty$, for y_2, \dots, y_n fixed and positive, this contradicts (1) since $a < \alpha_1$ and $\int_R |\phi|^2 d\omega_u > 0$. Thus we have a contradiction and hence $\phi \equiv 0$ outside C . Hence the two classes **P** and **E** are identical.

Next let f belong to the class **P** (\equiv **E**) and let us denote by K the intersection of all convex bodies in the u -space in whose exteriors $\phi \equiv 0$, and let $s(\lambda)$ be the supporting-function of K defined as in (2). Then f has the representation

$$(8) \quad f(z_1, \dots, z_n) = \left(\frac{1}{2\pi}\right)^{n/2} \int_K \phi(u) e^{-iu_1 z_1 - \dots - iz_n u_n} d\omega_u$$

and

$$(9) \quad \begin{aligned} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1 + i\lambda_1 \rho, \dots, x_n + i\lambda_n \rho)|^2 d\omega_x \\ = \int_K |\phi|^2 e^{2(\lambda_1 u_1 + \dots + \lambda_n u_n) \rho} d\omega_u \\ \leq e^{2s(\lambda)\rho} \int_K |\phi|^2 d\omega_u. \end{aligned}$$

Thus

$$(10) \quad \frac{1}{2} \limsup_{\rho \rightarrow \infty} \frac{1}{\rho} \log \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1 + i\lambda_1 \rho, \dots, x_n + i\lambda_n \rho)|^2 d\omega_x \leq s(\lambda).$$

In order to see that the actual limit in (10) exists and is equal to $s(\lambda)$ let us consider a fixed direction (λ^0) . Then there is an extreme point⁴ (u^0) of K such that $s(\lambda^0) = \lambda_1^0 u_1^0 + \dots + \lambda_n^0 u_n^0$. Moreover for $\delta > 0$ there clearly exists a neighborhood $N = N(\delta)$ of (u^0) such that

$$(11) \quad s(\lambda^0) - \delta \leq \lambda_1^0 u_1 + \dots + \lambda_n^0 u_n \leq s(\lambda^0) + \delta, \text{ for } (u) \in NK.$$

Hence

$$(12) \quad \begin{aligned} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1 + i\lambda_1^0 \rho, \dots, x_n + i\lambda_n^0 \rho)|^2 d\omega_x \\ = \int_K |\phi|^2 e^{2(\lambda_1^0 u_1 + \dots + \lambda_n^0 u_n) \rho} d\omega_u \geq \int_{NK} |\phi|^2 e^{2(\lambda_1^0 u_1 + \dots + \lambda_n^0 u_n) \rho} d\omega_u \\ \geq e^{2[s(\lambda^0) - \delta]\rho} \int_{NK} |\phi|^2 d\omega_u. \end{aligned}$$

⁴ By an extreme point of a convex body K is meant a boundary point which is not an inner point of any line segment of K . For each direction (λ) there is an extreme point which lies on the supporting plane in that direction, i. e. on $\lambda_1 u_1 + \dots + \lambda_n u_n = s(\lambda)$. An extreme point also possesses the property that if any neighborhood N of it is omitted from K , then the convex extension of $K - NK$ is a proper subset of K . For these properties see T. Bonnesen and W. Fenchel, "Theorie der Konvexen Körper," *Ergebnisse der Math. und ihrer Grenzgebiete*, Berlin (1934), esp. pp. 15, 16 or G. Pólya, "Untersuchungen über Lücken und Singularitäten von Potenzreihen," *Mathematische Zeitschrift*, vol. 29 (1929), pp. 549-640, esp. pp. 573-578.

Now $\int_{NK} |\phi|^2 d\omega_n > 0$ since otherwise ϕ would be identically zero in NK and thus ϕ would be identically zero in the exterior of the convex body K^* which is the convex extension of $K - NK$. But this is impossible since K^* is a proper subset of K (see ⁴) and this contradicts the definition of K . Hence (12) yields

$$(13) \quad \frac{1}{2} \liminf_{\rho \rightarrow \infty} \frac{1}{\rho} \log \int_{-\infty}^{\infty} \cdots \int |f(x_1 + i\lambda_1^0 \rho, \dots, x_n + i\lambda_n^0 \rho)|^2 d\omega_x \geq s(\lambda^0) - \delta.$$

From (10) and (13), since (λ^0) is an arbitrary direction and δ is an arbitrary positive number, it follows that the limit in (10) exists and that it is equal to $s(\lambda)$.

We have proved the following theorem.

THEOREM 1. *Let $f(z_1, \dots, z_n)$ be an entire function satisfying (1). Then the limit in (3) exists and is equal to $s(\lambda)$, where $s(\lambda)$ is the supporting function defined by (2) of the convex body K , where K is the intersection of all convex bodies in whose exteriors the Fourier transform of f is identically zero.*

Plancherel and Pólya (*loc. cit.*²) have considered the class **P** of functions and have shown that the growth function $h_P(\lambda)$ defined by them as in (4) is equal to the function $s(\lambda)$. Thus we have

COROLLARY. *If $f \in \mathbf{P}$ then*

$$(14) \quad \frac{1}{2} \lim_{\rho \rightarrow \infty} \frac{1}{\rho} \log \int_{-\infty}^{\infty} \cdots \int |f(x_1 + i\lambda_1 \rho, \dots, x_n + i\lambda_n \rho)|^2 d\omega_x \\ = \max_{\substack{-\infty < a_k < \infty \\ k=1, \dots, n}} \limsup_{\rho \rightarrow \infty} \frac{1}{\rho} \log |f(a_1 + i\lambda_1 \rho, \dots, a_n + i\lambda_n \rho)|.$$

In connection with Theorem 1, let us remark that Plancherel and Pólya (*loc. cit.*², p. 146) have shown that

$$\int_{-\infty}^{\infty} \cdots \int |f(x_1 + iy_1, \dots, x_n + iy_n)|^2 d\omega_x \\ \leq e^{2c(|y_1| + \dots + |y_n|)} \int_{-\infty}^{\infty} \cdots \int |f(x_1, \dots, x_n)|^2 d\omega_x$$

where c is the *cardinal increase* of f , that is, c is the greatest value of $h_P(\lambda)$ for (λ) ranging over all the sets for which one λ_k is ± 1 and all others are 0. They also obtain an analogous result for the class L^p .

2. Functions analytic in the "octant" $Q = E[I_m\{z_k\} > 0, k = 1, \dots, n]$. Let $f(z_1, \dots, z_n)$ be analytic in Q and let it satisfy a relation of the form (1) for all positive values of y_1, \dots, y_n . We shall obtain a result for this case analogous to that obtained in the previous section. Define

$$(15) \quad g(z_1, \dots, z_n) = e^{ia(z_1, \dots, z_n)} f(z_1, \dots, z_n).$$

Then by (1) (for y_1, \dots, y_n positive)

$$(16) \quad \begin{aligned} & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |g(x_1 + iy_1, \dots, x_n + iy_n)|^2 d\omega_x \\ &= e^{-2a(y_1 + \dots + y_n)} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1 + iy_1, \dots, x_n + iy_n)|^2 d\omega_x \\ &< A. \end{aligned}$$

Now Bergmann and Martin⁵ have shown that a function g analytic in Q and satisfying a relation of the form (16) for (y_1, \dots, y_n) positive has a Fourier transform $\gamma(u) e^{y_1 u_1 + \dots + y_n u_n}$ which vanishes outside the octant $q = E[u_k < 0, k = 1, \dots, n]$ and which has the property that $\gamma(u) \in L^2$. Thus g has a representation of the form

$$(17) \quad g(z_1, \dots, z_n) = \left(\frac{1}{2\pi}\right)^{n/2} \int_{-\infty}^0 \dots \int_{-\infty}^0 \gamma(u) e^{-iu_1 z_1 - \dots - iu_n z_n} d\omega_u, \quad (z) \in Q.$$

Using (15), we see that

$$(18) \quad \begin{aligned} f(z_1, \dots, z_n) &= \left(\frac{1}{2\pi}\right)^{n/2} \int_{-\infty}^0 \dots \int_{-\infty}^0 \gamma(u) e^{-i(u_1 + a)z_1 - \dots - i(u_n + a)z_n} d\omega_u \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \int_{-\infty}^a \dots \int_{-\infty}^a \phi(u) e^{-iu_1 z_1 - \dots - iu_n z_n} d\omega_u, \quad (z) \in Q, \end{aligned}$$

where

$$(19) \quad \phi(u_1, \dots, u_n) = \gamma(u_1 - a, \dots, u_n - a).$$

Thus the class of all functions f analytic in Q and satisfying relations of the form (1) for all positive values of y_1, \dots, y_n , is contained in the class of functions defined by

$$(20) \quad f(z_1, \dots, z_n) = \left(\frac{1}{2\pi}\right)^{n/2} \int_{-\infty}^a \dots \int_{-\infty}^a \phi(u) e^{-iu_1 z_1 - \dots - iu_n z_n} d\omega_u, \quad (z) \in Q,$$

where $\phi \in L^2$ over $-\infty < u_k \leq a, k = 1, \dots, n$, and vanishes identically elsewhere. That these two classes are identical follows at once. We omit the details.

⁵ S. Bergmann and W. T. Martin, "On a modified moment problem in two variables," to appear in the *Duke Mathematical Journal*. See esp. Theorem 1.

Next let S_p be a closed sphere of radius p , center $u_1 = \dots = u_n = 0$, and let K_p be the intersection of all convex bodies C in S_p such that $\phi \equiv 0$ in $S_p - C$. Then clearly $K_p \subset K_{p+1}$ and the point set K' consisting of all points in any K_p , $p = 1, 2, \dots$, has the property that $\phi \equiv 0$ outside K' . Thus

$$(21) \quad f(z_1, \dots, z_n) = \left(\frac{1}{2\pi}\right)^{n/2} \int_{K'} \phi(u) e^{-iu_1 z_1 - \dots - iu_n z_n} d\omega_u, \quad (z) \in Q,$$

and for positive $\lambda_1, \dots, \lambda_n$ we have

$$(22) \quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1 + i\lambda_1 \rho, \dots, x_n + i\lambda_n \rho)|^2 d\omega_x \\ = \int_{K'} |\phi|^2 e^{2(\lambda_1 u_1 + \dots + \lambda_n u_n) \rho} d\omega_u \leq e^{2s'(\lambda) \rho} \int_{K'} |\phi|^2 d\omega_u,$$

where

$$(23) \quad s'(\lambda) = \text{l. u. b. } \{\lambda_1 u_1 + \dots + \lambda_n u_n\}_{(u) \in K'}, \quad \lambda_1, \dots, \lambda_n \text{ positive.}^6$$

Hence

$$(24) \quad \frac{1}{2} \limsup_{\rho \rightarrow \infty} \frac{1}{\rho} \log \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1 + i\lambda_1 \rho, \dots, x_n + i\lambda_n \rho)|^2 d\omega_x \leq s'(\lambda)$$

for positive λ 's. Again we can show that the actual limit exists and that it is equal to $s'(\lambda)$. For this purpose let us apply Plancherel's theorem to (21). Then

$$(25) \quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1 + i\lambda_1 \rho, \dots, x_n + i\lambda_n \rho)|^2 d\omega_x \\ \geq \int_{K_p} |\phi|^2 e^{2(\lambda_1 u_1 + \dots + \lambda_n u_n) \rho} d\omega_u, \quad (p = 1, 2, \dots).$$

Now by Theorem 1 we have

$$(26) \quad \frac{1}{2} \lim_{\rho \rightarrow \infty} \frac{1}{\rho} \log \int_{K_p} |\phi|^2 e^{2(\lambda_1 u_1 + \dots + \lambda_n u_n) \rho} d\omega_u = s_p(\lambda)$$

and thus

$$(27) \quad \frac{1}{2} \liminf_{\rho \rightarrow \infty} \frac{1}{\rho} \log \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1 + i\lambda_1 \rho, \dots, x_n + i\lambda_n \rho)|^2 d\omega_x \geq s_p(\lambda), \\ (p = 1, 2, \dots),$$

for $\lambda_1, \dots, \lambda_n$ positive. This implies that the left-hand side of (27) is greater than or equal to $s'(\lambda)$ for positive λ 's. For let (λ^0) be a positive

⁶ It is clear that $K' \subset E[-\infty < u_k < a, K = 1, \dots, n]$ and hence that
 $\text{l. u. b. } \{\lambda_1 u_1 + \dots + \lambda_n u_n\} \leq (\lambda_1 + \dots + \lambda_n) a$ for $\lambda_1, \dots, \lambda_n$ positive.
 $(u) \in K'$

direction. Then in view of the definition (23) of $s'(\lambda)$ there is a sequence (u^p) of points such that $(u^p) \in K_{v_p}$, (where $v_p \rightarrow \infty$ as $p \rightarrow \infty$) and such that

$$(28) \quad \lambda_1^0 u_1^p + \cdots + \lambda_n^0 u_n^p \rightarrow s'(\lambda^0) \quad \text{as } p \rightarrow \infty.$$

The relation (28) together with the fact that $\lambda_1^0 u_1^p + \cdots + \lambda_n^0 u_n^p \leq s_{v_p}(\lambda^0)$ gives

$$\liminf_{p \rightarrow \infty} s_{v_p}(\lambda^0) \geq s'(\lambda^0),$$

and hence the left-hand side of (27) (for $(\lambda) = (\lambda^0)$) is greater than or equal to $s'(\lambda^0)$. Since (λ^0) is an arbitrary positive direction this result together with (24) yields the following result.

THEOREM 2. *If f is analytic in Q and if (1) holds for positive y_1, \dots, y_n then*

$$\frac{1}{2} \lim_{\rho \rightarrow \infty} \frac{1}{\rho} \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |f(x_1 + i\lambda_1 \rho, \dots, x_n + i\lambda_n \rho)|^2 d\omega_x = s'(\lambda)$$

for positive λ 's, where $s'(\lambda)$ is defined as in equation (23).

THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

PROJECTIVE ANALOGUES OF THE CONGRUENCE OF NORMALS.*

By PHILIP O. BELL.

1. Introduction. The *projective normal* at a point of a non-ruled surface S in ordinary space was defined by Fubini as the cusp-axis of y with respect to the extremal curves of his integral invariant

$$\int (a'bv')^{\frac{1}{2}} du.$$

It is well known that the *pseudo-normal* which Green proposed as a projective analogue of the normal coincides with the projective-normal.

Green and Fubini discovered, quite independently, certain analogies which exist between this line and the normal. Green noted that the projective normal, like the normal, is intrinsically connected with the surface, and that the curves which correspond to the developables of the projective normal congruence resemble the lines of curvature by also forming a *conjugate net*. Fubini's considerations reveal that both normal and projective normal may be defined as cusp-axes of certain integral invariants. Fubini's definition lacks geometric significance without a geometric interpretation for his integral invariant. The author [1, p. 403]¹ has recently provided such an interpretation.

Green designated a congruence whose developables correspond to a conjugate net a *conjugate congruence*. Grove [2] has proved analytically the existence of a class of covariant conjugate congruences a general one of which he calls an *R-conjugate congruence*. He does not however characterize geometrically any one of these congruences. It is the purpose of this paper to present a method for the geometric determination of a general *R-conjugate congruence* and to show that it is also characterized by the other important property of the projective normal by being similarly determined with respect to the extremals of an integral invariant. A method will also be given for the geometric interpretation of these extremals. Finally certain special *R-conjugate congruences* will be introduced.

Let the surface S be referred to its asymptotic net as parametric, with the fundamental differential equations in Wilczynski's canonical form

$$(1.1) \quad y_{uu} + 2by_v + fy = 0, \quad y_{vv} + 2a'y_u + gy = 0.$$

* Received November 10, 1939.

¹ Numbers in brackets refer to the bibliography at the end of the paper.

Using the notation introduced in the celebrated memoir by Green [3] let us consider the parametric vector equations

$$(1.2) \quad y = y(u, v), \quad \rho = y_u - \beta y, \quad \sigma = y_v - \alpha y, \quad \tau = y_{uv} - \alpha y_u - \beta y_v + \alpha \beta y$$

where α and β are arbitrary analytic functions of u and v . Equations (1.2) define the general homogeneous coördinates of four points which we denote simply by y, ρ, σ and τ when no confusion can arise. The line l joining the points ρ, σ , according to Green's classification, is an arbitrary line of the first kind and generates a congruence Γ of the first kind as y moves over S . The reciprocal l' of the line l with respect to S at y is an arbitrary line of the second kind and generates a congruence Γ' of the second kind as y moves over S . If the functions α, β are chosen suitably the points y, ρ, σ and τ become covariant points and the congruences Γ and Γ' become covariant congruences.

2. Conjugate congruences. Consider any two covariant points ω_1 and ω_2 which are collinear with y but do not lie in the tangent plane to S at y . The general coördinates for ω_1 and ω_2 are given by $\omega_i = \tau + r_i y$, ($i = 1, 2$), where τ is defined by (1.2) and r_1 and r_2 are functions of u, v . The tangent lines at ω_1 and ω_2 to the curves described by these points as y moves along a curve C_λ , defined by $dv - \lambda(u, v)du = 0$, intersect the tangent plane to S at y in the points which we denote by $W_1^{(\lambda)}$ and $W_2^{(\lambda)}$. Expressions for the coördinates of $W_i^{(\lambda)}$, ($i = 1, 2$) are linear combinations of ω_i and $(\omega_i)_u + \lambda(\omega_i)_v$ which do not contain y_{uv} . The terms of $(\omega_i)_u + \lambda(\omega_i)_v$ which involve y_{uv} are equal to $-(\beta + \alpha\lambda)y_{uv}$. Hence, the expressions for the general coördinates of $W_1^{(\lambda)}$ and $W_2^{(\lambda)}$ are given by

$$(2.1) \quad W_i^{(\lambda)} = (\omega_i)_u + \lambda(\omega_i)_v + (\beta + \alpha\lambda)\omega_i, \quad (i = 1, 2).$$

Making use of the forms for ω_1 and ω_2 and the equations (2.1), we have

$$(2.2) \quad W_2^{(\lambda)} - W_1^{(\lambda)} = R[(y_u - \beta y) + \lambda(y_v - \bar{\alpha}y)],$$

where

$$\bar{\beta} = -(\beta + [\log R]_u), \quad \bar{\alpha} = -(\alpha + [\log R]_v), \quad R = r_2 - r_1.$$

Let t_λ denote the tangent to C_λ at y . Let ν_λ denote the point of intersection of t_λ and the line joining $W_1^{(\lambda)}$ and $W_2^{(\lambda)}$. The right hand member of (2.2) is clearly the expression for the general coördinates of ν_λ . We shall call the point ν_λ the ν -point of t_λ , corresponding to the points ω_1 and ω_2 .

Since the right hand member of (2.2) is a linear combination of $y_u - \beta y$ and $y_v - \bar{\alpha}y$, the point ν_λ , for any value of λ , lies on a straight line \bar{l} which joins $\bar{\rho}$ and $\bar{\sigma}$ given by

$$\bar{\rho} = y_u - \bar{\beta}y, \quad \sigma = y_v - \bar{\alpha}y,$$

where $\bar{\beta}$ and $\bar{\alpha}$ are defined above. Hence, we have the theorem

THEOREM (2.1). *As the direction λ is varied, while u and v are held constant, the v -point of t_λ , corresponding to the points ω_1 and ω_2 describes a straight line \bar{l} .*

The point μ of intersection of the line \bar{l} with the reciprocal of the line joining ω_1 and ω_2 has general coördinates of the form

$$\mu = (2\alpha + [\log R]_v)(y_u + [\log R]_u y/2) - (2\beta + [\log R]_u)(y_v + [\log R]_v y/2).$$

Let t_μ denote the tangent to S at y which passes through the point μ . In view of the forms of the functions $\bar{\alpha}$, $\bar{\beta}$, we have

THEOREM (2.2). *The harmonic conjugate of the tangent t_μ with respect to the line \bar{l} and the reciprocal of the line joining ω_1 and ω_2 is the R -harmonic line, which joins the points ρ and σ given by $\rho = y_u + (\log R)_u y/2$ and $\sigma = y_v + (\log R)_v y/2$. The reciprocal of this line is the R -conjugate line.*

To complete the characterization of the R -conjugate line for a given function $R = R(u, v)$ it is, of course, necessary to have geometric definitions of covariant points ω_1 and ω_2 whose general coördinates are related by the equation $\omega_2 = \omega_1 + kRy$, $k = \text{const.}$

The integral $\int (Rv')^{1/2} du$, where $R(u, v)$ is associated with covariant points ω_1, ω_2 in the manner described in the preceding paragraph, is an integral invariant which is projectively and intrinsically related to the arc of a curve along which it is calculated. The extremals of this integral are defined by the curvilinear equation

$$(2.3) \quad v'' = (\log R)_u v' - (\log R)_v v'^2.$$

It is well known that if Wilczynski's canonical form (1.1) is used, the cusp-axis of y with respect to a two parameter family of hypergeodesics defined by

$$v'' = A + Bv' + Cv'^2 + Dv'^3,$$

passes through y and the point z given by $z = y_{uv} - \alpha y_u - \beta y_v$, where $\alpha = C/2$, $\beta = -B/2$. Hence, we have

THEOREM (2.3). *The R -conjugate line is the cusp-axis of y with respect to the extremal curves of the integral invariant $\int (Rv')^{1/2} du$.*

To add to the geometric significance of the above theorem the extremal curves of the integral $\int (Rv')^{1/2} du$ will be geometrically characterized.

THEOREM (2.4). *The tangent t_μ associated geometrically with covariant points ω_1 and ω_2 which lie on the cusp-axis of y with respect to a pencil p_λ of conjugate nets and the tangent t_λ of the curve C_λ of the fundamental net N_λ at y are conjugate tangents if, and only if, the curve C_λ is an extremal of the integral invariant $\int (Rv')^{1/2} du$, where $kRy = \omega_2 - \omega_1$, $k = \text{const}$.*

According to the hypothesis we must have

$$(2.4) \quad \lambda = (2\beta + [\log R]_u) / (2\alpha + [\log R]_v),$$

where $\beta = -(\log \lambda)_u / 2$ and $\alpha = (\log \lambda)_v / 2$. Hence, on clearing of fractions we obtain

$$(2.5) \quad \lambda_u + \lambda\lambda_v = (\log R)_u \lambda - (\log R)_v \lambda^2,$$

which, on substituting v' for λ and v'' for $\lambda_u + \lambda\lambda_v$, becomes equation (2.3). The operations are reversible and therefore the condition is necessary and sufficient.

3. Special conjugate congruences. The projective normal is the special case of the R -conjugate line for which $R = ka'b$, $k = \text{arbitrary const}$. To complete its geometric characterization it is necessary to locate two points ω_1 and ω_2 whose general coördinates are related by the equation $\omega_2 - \omega_1 = ka'by$, $k = \text{const}$. Two such points are the intersections (distinct from y) of an arbitrary line l' of the second kind with the quadric of Wilczynski and the quadric of Lie.

From the standpoint of analytic simplicity the projective normal is the best available projective substitute for the normal. From a geometric point of view, however, it is quite conceivable that there may be other R -conjugate lines equally suitable as a projective substitute for the normal. An R -conjugate line of this character will be introduced in connection with a new pencil of quadric surfaces.

Let l_k denote a general line of the first canonical pencil. The line l_k intersects the u and v -tangents to S at y in the points ρ , σ defined in (1.2), where

$$(3.1) \quad \begin{cases} \beta = (\log a'^2b)_u / k - (\log a'b)_u / 2, \\ \alpha = (\log b^2a')_v / k - (\log a'b)_v / 2. \end{cases}$$

The lines l_2 , l_3 , l_4 and l_∞ are the first directrix of Wilczynski, the reciprocal of the axis of Čech, the first canonical edge of Green, and the reciprocal of the projective normal, respectively. As y moves over S the points ρ , σ of l_k generate transversal surfaces S_ρ and S_σ of the congruence described by l_k . The v -tangent at ρ to S_ρ intersects l'_k , the reciprocal of l_k , in the point which

we denote by η_k whose coördinates are given by $\eta_k = \tau - \beta_v y$, where α, β are the functions associated with l_k . Likewise the u -tangent at σ to S_σ intersects l'_k in the point which we denote by ξ_k whose coördinates are given by $\xi_k = \tau - \alpha_u y$.² Let ζ_k denote the harmonic conjugate of y with respect to the points η_k and ξ_k . The general coördinates of ζ_k may be easily found to be given by $\zeta_k = \tau + (k-3)(\log a'b)_{uv}y/2k$, where the functions α, β in the expression for τ are given by (3.1). It is well known that just one quadric of Darboux at y passes through a given point not in the tangent plane to S at y . The equation of the unique quadric of Darboux which passes through the point ζ_k , $k = \text{const.}$, is easily found to be

$$(3.2) \quad x_2x_3 - x_1x_4 + (k-3)(\log a'b)_{uv}x_4^2/2k = 0.$$

This quadric is, therefore, a general member of a pencil of quadrics whose members are in one to one correspondence with the lines of the first canonical pencil. The quadrics of this pencil will therefore be called *canonical quadrics*. The special case of (3.2) for $k=3$ is clearly the *canonical quadric of Wilczynski*. Stouffer [4], without introducing the general quadric (3.2), has given the above characterization for the quadric of Wilczynski.

The intersection of a general line l' of the second kind with the quadric (3.2) is a point, which we denote by ω_k , whose general coördinates are given by $\omega_k = \tau + (k-3)(\log a'b)_{uv}y/2k$, where the functions α, β in the expression for τ are arbitrary. The form for the coördinates of ω_k shows that $\omega_j - \omega_k = c(\log a'b)_{uv}y$ where $c = (j-3)/2j - (k-3)/2k$, $j, k = \text{const.'s}$. Hence, the following theorem is an immediate consequence.

THEOREM (3.1). *If the fundamental points ω_1 and ω_2 are chosen as the intersection of a line l' of the second kind with two quadrics from the pencil of canonical quadrics, the associated R -conjugate line is independent of the choice of l' and is independent of the selection of the two quadrics of the pencil. For this line the associated functions α, β are given by $\alpha = -(\log R)_v/2$, $\beta = -(\log R)_u/2$, where $R = (\log a'b)_{uv}$.*

4. R -conjugate congruences associated with one-parameter families of curves.

The transformation

$$(4.1) \quad y = x/(R)^{1/2}$$

² The points η_k and ξ_k are special cases of the points η_1 and η_2 , respectively, which were introduced by Green [3, p. 95].

transforms the covariant points $(R)^{1/2}(y_u + R_u y/2R)$, $(R)^{1/2}(y_v + R_v y/2R)$ and

$$(R)^{1/2}\{y_{uv} + R_v y_u/2R + R_u y_v/2R + [R_u R_v/4R^2 + (\log R)_{uv}/2]y\}$$

into x_u , x_v and x_{uv} respectively. The points x_u , x_v are the intersections of the R -harmonic line with the asymptotic u and v -tangents to S at x , and the point x_{uv} lies on the R -conjugate line and is characterized like a point ξ_k , but with l_k replaced by the R -harmonic line. The effect of transformation (4.1) on system (1.1) is to produce the following *canonical form*

$$(4.2) \quad \begin{cases} x_{uu} = px + \theta_u x_u + \beta x_v, \\ x_{vv} = qx + \gamma x_u + \theta_v x_v, \end{cases}$$

wherein,

$$\begin{aligned} \theta &= \log R, & \beta &= -2b, & \gamma &= -2a' \\ p &= -f + b\theta_v + \theta_{uu}/2 - \theta_u^2/4 & \text{and} & & q &= -g + a'\theta_u + \theta_{vv}/2 - \theta_v^2/4. \end{aligned}$$

If $R = a'b$, the form (4.2) is *Fubini's canonical form*.

The intersection of the R -harmonic line with the tangent at x to the curve C_λ defined by $dv - \lambda du = 0$ is the point $x_u + \lambda x_v$. The tangent plane at $x_u + \lambda x_v$ to the ruled surface described by the R -harmonic line as x moves along C_λ intersects the R -conjugate line in a point P_λ whose general coördinates are found to be given by

$$(4.3) \quad P_\lambda = x_{uv} + (p + q\lambda^2)x/2\lambda.$$

The Γ -curves of the R -harmonic congruence form a conjugate net N_{λ_1} whose curvilinear differential equation is

$$(4.4) \quad dv^2 - \lambda_1^2 du^2 = 0, \quad \text{where} \quad \lambda_1 = (p/q)^{1/2}.$$

The points $P_{-\lambda_1}$, P_{λ_1} associated with the curves of N_{λ_1} which pass through the point x are given by

$$P_{-\lambda_1} = x_{uv} - (pq)^{1/2}x, \quad P_{\lambda_1} = x_{uv} + (pq)^{1/2}x.$$

We recall that a conjugate line may be determined in association with an arbitrary line l' by choosing fundamental points ω_1 , ω_2 on l' and following the method outlined in § 2. The conjugate line thus determined with respect to the fundamental points $P_{-\lambda_1}$, P_{λ_1} which lie on an arbitrary chosen R -conjugate line is especially interesting because of its remarkable analytic, as well as geometric, simplicity. By making use of equations (4.2) in carrying out the analysis for this determination we obtain the following

THEOREM (4.1). *The R -conjugate line ($R = [pq]^{1/2}$) determined with respect to the points $P_{-\lambda_1} = x_{uv} - (pq)^{1/2}x$, $P_{\lambda_1} = x_{uv} + (pq)^{1/2}x$, of an arbitrary line l' is the line l' .*

trarily chosen R -conjugate line, as fundamental, passes through the points x and $z = x_{uv} - ax_u - bx_v$ where a, b are defined by

$$a = [\log(R/(pq)^{1/2})]_v/2, \quad b = [\log(R/(pq)^{1/2})]_u/2.$$

This line is the cusp-axis of the point x with respect to the extremal curves of the integral invariant

$$\int (pq)^{1/4} v'^{1/2} du.$$

Of course, other R -conjugate congruences may be associated with a given one by selecting points $P_{z\lambda}$ which are associated with other significant curves of S . The investigation of some of these may prove interesting.

UNIVERSITY OF KANSAS.

BIBLIOGRAPHY.

-
- (1) P. O. Bell, "A study of curved surfaces by means of certain associated ruled surfaces," *Transactions of the American Mathematical Society*, vol. 46 (1939), pp. 389-409.
 - (2) V. G. Grove, "On canonical forms of differential equations," *Bulletin of the American Mathematical Society*, vol. 36 (1930), pp. 582-586.
 - (3) G. M. Green, "Memoir on the general theory of surfaces and rectilinear congruences," *Transactions of the American Mathematical Society*, vol. 20 (1919), pp. 79-153.
 - (4) E. B. Stouffer, "A geometrical determination of the canonical quadric of Wilczynski," *Proceedings of the National Academy of Sciences* (18), vol. 3 (1932), pp. 252-255.

CONVERGENCE THEOREMS FOR FUNCTIONS OF TWO COMPLEX VARIABLES.*

By WILLIAM F. WHITMORE.

1. Introduction. The theory of harmonic measure has proved a very valuable tool in the theory of functions of one complex variable. The possibility of these applications is due on the one hand to the fact that the real or imaginary part of an a. f. 1 c. v. (analytic function of one complex variable) is a harmonic function and on the other to the fact that the Dirichlet problem can be solved uniquely in terms of harmonic functions, thus assuring the existence of the harmonic measure.¹ In attempting to carry over these ideas to functions of two complex variables, one is confronted by the fact that it is not possible to prescribe arbitrary boundary values for a biharmonic function (real or imaginary part of an a. f. 2 c. v.) on the entire three dimensional boundary of a four dimensional domain. In order to preserve at least a portion of the properties of the one variable case, Bergmann (B_1) has introduced the concept of domains with distinguished boundary surface. The three dimensional boundary of such a domain contains a closed two dimensional manifold—the *distinguished surface* (ausgezeichnete Randfläche, surface remarquable)—which has properties for the theory of a. f. 2 c. v. analogous to those of the boundary for the one variable case, in that a regular a. f. 2 c. v.

* Received September 21, 1939.

¹ The method of approach used here has been chiefly developed by Stefan Bergmann in a long series of papers, of which I have had occasion to cite five in particular:

- (B_1) "Ueber die ausgezeichneten Randflächen in der Theorie der Funktionen von zwei komplexen Veränderlichen," *Mathematische Annalen*, vol. 104 (1931), pp. 611-636.
- (B_2) "Zwei Sätze aus dem Ideenkreis des Schwarzsehen Lemma bei den Funktionen von zwei komplexen Veränderlichen," *Mathematische Annalen*, vol. 109 (1934), pp. 324-348.
- (B_3) "Ueber eine Integraldarstellung von Funktionen zweier komplexer Veränderlichen," *Mathematicheskii Sbornik*, vol. 1 (43) (new series), pp. 851-861.
- (B_4) "Ueber eine in gewissen Bereichen mit Maximumfläche gültige Integraldarstellung der Funktionen zweier Variabler," *Mathematische Zeitschrift*, vol. 39, pp. 605-608.
- (B_5) "Ueber eine Abschätzung von meromorphen Funktionen zweier komplexer Veränderlichen in Bereichen mit ausgezeichneter Randfläche," *Travaux de l'Inst. Math. Tbilissi*, vol. 1, pp. 187-204.

The theory of harmonic measure for one variable is given in Nevanlinna: "Eindeutige Analytische Funktionen," here cited as (N).

attains its maximum on this surface, a biharmonic function is determined by its values there, etc. An example of a domain with distinguished surface is given by any domain bounded by a finite number of analytic hypersurfaces (three dimensional manifolds defined by analytic relations between the two complex variables), the distinguished surface being formed by the intersections of these hypersurfaces—e. g., the bicylinder $|z_1| < 1, |z_2| < 1$ with a boundary composed of the two analytic hypersurfaces $z_1 - e^{i\theta_1} = 0, |z_2| < 1$ and $z_2 - e^{i\theta_2} = 0, |z_1| < 1$ has the distinguished surface $|z_1| = 1, |z_2| = 1$.

Although a biharmonic function is uniquely determined by its values on the distinguished surface of a domain, it is in general not possible to find a biharmonic function defined in the domain which assumes arbitrarily prescribed values on this surface. Hence a biharmonic measure cannot be used to generalize the notion of harmonic measure, for such a measure may not exist. Bergmann (B_2) has met this further complexity by introducing the notion of *functions of extended class*. This class possesses properties necessary for the extension of harmonic measure; in particular, the property that to every bounded, piecewise continuous function given on the distinguished surface of a domain there corresponds a unique function of the extended class defined in the domain, and also that the operator defining the class is linear. The class depends, in general, on the domain. For a domain where the range of each complex variable is independent of the other (product domain, also called cylinder domain), the extended class is known to be the class of doubly harmonic functions (B_2), so that for such domains the notion of harmonic measure can be replaced by that of *doubly harmonic measure*. For functions of 1 c. v., Lindelöf has proved a theorem to the following effect (N, p. 44):

If an analytic function defined and bounded in the upper half-plane converges to a limit for z tending to infinity along the negative real axis, then it converges to the same value uniformly in each angle-space $\pi > \arg z > \eta > 0$.

Or stated for the unit circle:

If an analytic function defined and bounded in the unit circle converges to a limit for one-sided approach along the boundary to a given boundary point, then it converges to the same value uniformly along any path in the interior which ends at the given point and makes a positive angle with the circumference at the point.

With the aid of the theory of doubly harmonic measure we shall show that analogous results can be established on the convergence of bounded functions of 2 c. v. defined in certain domains with distinguished surface.

2. Notation and definitions. We consider functions of the two complex variables $z_k = x_k + iy_k$ ($k = 1, 2$). A doubly harmonic function of the four real variables x_1, y_1, x_2, y_2 is defined by the equations

$$(1) \quad \frac{\partial^2 u}{\partial x_k^2} + \frac{\partial^2 u}{\partial y_k^2} = 0 \quad (k = 1, 2).$$

A biharmonic function is the real or imaginary part of an a. f. 2 c. v. and satisfies in addition to equations (1) the equations

$$(2) \quad \frac{\partial^2 u}{\partial x_1 \partial x_2} + \frac{\partial^2 u}{\partial y_1 \partial y_2} = 0; \quad \frac{\partial^2 u}{\partial x_1 \partial y_2} - \frac{\partial^2 u}{\partial x_2 \partial y_1} = 0,$$

as can be verified by application of the Cauchy-Riemann equations. The symbol \cdot indicates the intersection of two point sets; the symbol \times indicates their topological product. $E[\cdot \cdot \cdot]$ denotes the set of points satisfying the relations enclosed in the brackets. A four or two dimensional domain will be indicated by a capital letter and the corresponding three or one dimensional boundary by the corresponding small letter. An upper index j attached to the symbol for a set gives its dimensionality ($0 < j < 4$); e. g., \mathfrak{F}^2 is a two dimensional set. Let \mathcal{G}_k^2 ($k = 1, 2$) be a domain in the z_k -plane, bounded by a finite number of Jordan arcs g_k^1 (\mathcal{G}_k^2 may be multiply connected). The product domain $\mathfrak{A} = \mathcal{G}_1^2 \times \mathcal{G}_2^2$ is a four dimensional domain in the (z_1, z_2) -space. The two dimensional surface $\mathfrak{F}^2 = g_1^1 \times g_2^1$ is the distinguished surface of \mathfrak{A} . As noted in the introduction, the Dirichlet problem of determining a function defined in \mathfrak{A} which assumes prescribed bounded and piecewise continuous values on \mathfrak{F}^2 can be solved uniquely in terms of doubly harmonic functions; in the case of a bicylinder, an explicit form for the desired function is given by an iterated Poisson integral (B_2). Hence, if \mathfrak{F}^2 is a subset of \mathfrak{F}^2 having positive two dimensional measure, there exists a unique doubly harmonic function defined in \mathfrak{A} which assumes the value 1 on \mathfrak{F}^2 and the value 0 on $\mathfrak{F}^2 - \mathfrak{F}^2$. This function will be denoted by $\omega(z_1, z_2; \mathfrak{F}^2, \mathfrak{A})$ and is defined to be the doubly harmonic measure of \mathfrak{F}^2 with respect to \mathfrak{A} taken at the point $\{z_1, z_2\}$.

3. Convergence in Bicylinders. Using the notion of functions of extended class, Bergmann has established for domains with distinguished surface a generalization of a theorem given by Ostrowski for the one variable case (B_2 , p. 344). It will be stated here in the restricted case of product domains with the aid of doubly harmonic measure.²

² Note that in Bergmann's statement a summation sign is omitted.

THEOREM 1. Let $f(z_1, z_2)$ be an a. f. 2 c. v. defined and regular in a product domain \mathfrak{A} and continuous on the boundary \mathfrak{A}^3 of \mathfrak{A} . Let the distinguished surface \mathfrak{F}^2 of \mathfrak{A} be composed of m disjunct pieces, $\mathfrak{F}^2 = \sum_{k=1}^m \mathfrak{F}_k^2$; and let $\omega(z_1, z_2; \mathfrak{F}_k^2, \mathfrak{A})$ be the doubly harmonic measure of \mathfrak{F}_k^2 . If there exist m constants M_k ($k=1, \dots, m$) such that $|f(z_1, z_2)| \leq M_k$ for $\{z_1, z_2\} \in \mathfrak{F}_k$, then one has in \mathfrak{A} the inequality:

$$(3) \quad \log |f(z_1, z_2)| \leq \sum_{k=1}^m (\log M_k) \omega(z_1, z_2; \mathfrak{F}_k^2, \mathfrak{A}).$$

For the case $m=2$ this theorem becomes a generalization of the so-called "two-constant" theorem (N, p. 41):

THEOREM 2. Let $f(z_1, z_2)$ and \mathfrak{A} satisfy the hypotheses of Theorem 1. If $|f(z_1, z_2)| \leq M$ on \mathfrak{F}^2 and $|f(z_1, z_2)| \leq m$ ($m < M$) on a subset \mathfrak{F}^2 of \mathfrak{F}^2 , then

$$(4) \quad \log |f(z_1, z_2)| < \mu \log m + (1 - \mu) \log M$$

at all points of the set $\{z_1, z_2\} \in E[\omega(z_1, z_2; \mathfrak{F}^2, \mathfrak{A}) > \mu, 1 \geq \mu \geq 0]$.

With the aid of Th. 2, the first of the desired convergence theorems can be established.

THEOREM 3. Given $f(z_1, z_2)$ an a. f. 2 c. v. defined and regular in the closed quarter-space $y_1 \geq 0, y_2 \geq 0$ (topological product of two upper half-planes), with $|f(z_1, z_2)| \leq 1$ on the distinguished surface $y_1 = y_2 = 0$. If $|f(z_1, z_2)| < \epsilon$ ($0 < \epsilon < 1$) for $\{z_1, z_2\} \in E[y_1 = y_2 = 0, \delta(x_1 + a) + x_2^2 < 0]$ where δ and a are arbitrary positive constants, then $|f(z_1, z_2)| < \epsilon^\mu$ for

$$(5) \quad \{z_1, z_2\} \in E[\arg(\delta(z_1 + a) + z_2^2) > \mu\pi, 1 > \mu > 0].$$

Proof. Apply Theorem 2 with $m = \epsilon, M = 1$. The function

$$\frac{1}{\pi} \arg(\delta(z_1 + a) + z_2^2) = \frac{1}{\pi} \arctan \frac{\delta y_1 + 2x_2 y_2}{\delta(x_1 + a) + x_2^2 - y_2^2}$$

is a doubly harmonic (in fact, biharmonic) function which is 1 on

$$\mathfrak{F}^2 = E[y_1 = y_2 = 0, \delta(x_1 + a) + x_2^2 < 0]$$

and 0 on the remainder of \mathfrak{F}^2 and hence is the doubly harmonic measure of \mathfrak{F}^2 .

Two remarks can be made concerning this result. First, the proof can obviously be extended without change to the case where z_2^2 and x_2^2 are replaced by z_2^{2n} and x_2^{2n} respectively. Second, in the limiting case where δ is allowed to approach infinity, the parabola $\delta(x_1 + a) + x_2^2 = 0$ becomes the line

$x_1 = -a$, and the theorem reduces to the ordinary one variable result, the convergence being supposed to depend only on the variable z_1 .

Put in another form, Theorem 3 says that if a bounded function converges to zero for approach to infinity in the real plane in such fashion that to every ϵ there corresponds a δ and an a so that $|f(z_1, z_2)| < \epsilon$ for every $\{z_1, z_2\}$ belonging to the set \mathfrak{Z}^2 of the theorem, then $|f(z_1, z_2)| < \epsilon^\mu$ (μ a fixed positive quantity less than 1) throughout the four dimensional domain (5), depending only on the parameters δ and a . Thus, if $\lim_{a \rightarrow \infty} \epsilon = 0$, convergence to zero uniformly in \mathfrak{Z}^2 implies convergence to zero in the domain (5) also.

The pair of linear transformations

$$(6) \quad \xi_1 = \frac{a(1 - e^{i\phi(a)})(z_1 + r)}{(1 + e^{i\phi(a)})(z_1 - r)}; \quad \xi_2 = i \frac{1 + z_2}{1 - z_2}$$

map the quarter-space $Im \xi_1 \geq 0, Im \xi_2 \geq 0$ on the bicylinder $|z_1| \leq r, |z_2| \leq 1$. The first transformation takes the points $(\infty, -a, 0)$ in the ξ_1 -plane into the points $(r, re^{i\phi(a)}, -r)$ in the z_1 -plane, so that the segment $(-\infty, -a)$ goes into the arc $(r, re^{i\phi(a)})$; here $\phi(a)$ is any suitable function of a for which $\phi(a) \rightarrow 0$ as $a \rightarrow \infty$. The second transformation takes $(0, 1, \infty)$ in the ξ_2 -plane into $(-1, -i, 1)$ in the z_2 -plane. Since such a mapping of the quarter-space on a bi-cylinder leaves the doubly harmonic measure invariant (the Poisson integral is invariant under linear transformations), it can be applied to the domain employed in Theorem 3 to give a convergence theorem for a bicylinder. The first transformation takes $Re(\xi_1 + a)$ into

$$\frac{2a[(1 + \cos \phi)|z_1 - r|^2 - 2y_1 r \sin \phi]}{|1 + e^{i\phi(a)}|^2 |z_1 - r|^2}$$

and the second takes $Re \xi_2$ into $\frac{-2y_2}{|1 - z_2|^2}$, so that Theorem 3 becomes:

THEOREM 3a. *Given $f(z_1, z_2)$ an a. f. 2 c. v. defined and regular in the closed bicylinder $|z_1| \leq r, |z_2| \leq 1$, with $|f(z_1, z_2)| \leq 1$ on the distinguished surface $|z_1| = r, |z_2| = 1$. If $|f(z_1, z_2)| < \epsilon$ ($0 < \epsilon < 1$) for*

$$(7) \quad \{z_1, z_2\} \in E \left[\frac{2a\delta[(1 + \cos \phi)|z_1 - r|^2 - 2y_1 r \sin \phi]}{|1 + e^{i\phi(a)}|^2 |z_1 - r|^2} + \frac{4y_2^2}{|1 - z_2|^4} < 0, |z_1| = r, |z_2| = 1 \right],$$

where δ and a are arbitrary positive constants, then $|f(z_1, z_2)| < \epsilon^\mu$ for

$$(8) \quad \{z_1, z_2\} \in E \left[\arg \left(\frac{2a\delta(z_1 - re^{i\phi(a)})}{(1 + e^{i\phi(a)})(z_1 - r)} - \left(\frac{1 + z_2}{1 - z_2} \right)^2 \right) > \mu\pi, 0 \leq \mu \leq 1 \right],$$

where $\phi(a)$ ($0 < \phi < \pi/2$) is any suitable function of a which tends to zero as a tends to infinity.

4. The \mathfrak{M} -domain and its properties. An \mathfrak{M} -domain is a four dimensional domain defined by

$$(9) \quad \mathfrak{M} = E[z_1 = th(z_2, \lambda), |z_2| < 1, 0 \leq t < 1, 0 \leq \lambda \leq 2\pi, \\ h(z_2, 0) = h(z_2, 2\pi)].$$

Its distinguished surface is $E[z_1 = h(z_2, \lambda), |z_2| = 1]$. The function $h(z_2, \lambda)$ is subject to the following conditions:

- (a) $h(z_2, \lambda)$ is an analytic function of z_2 ;
- (b) $h(z_2, \lambda)$ is a continuous function of λ whose derivative with respect to λ exists and is finite;
- (c) each curve $z_1 = h(z_2^0, \lambda)$ ($z_2^0 = \text{const.}$, $0 \leq \lambda \leq 2\pi$) has a positive radius of curvature at all points, the limit inferior of all radii of curvature being positive, and is such that any sufficiently small arc α^1 lies entirely within a circle whose circumference cuts the curve only at the endpoints of α^1 , whose radius is not less than the distance between these end-points, and whose center lies in the interior of the curve.³

It is possible to extend to \mathfrak{M} -domains a theorem on analytic continuation needed in what follows; the result was first established by Hartogs⁴ in the case of product domains.

LEMMA 1. *Given an \mathfrak{M} -domain defined by (9), assume that \mathfrak{M} contains a product domain $\mathfrak{R}^2 \times E[|z_2| < 1]$ (\mathfrak{R}^2 simply connected). Let $f(z_1, z_2)$ be a function satisfying the following conditions:*

- (a) $f(z_1, z_2)$ is an a. f. 2 c. v. in the interior of the product domain; and if z_1^0 is any given interior point of \mathfrak{R}^2 , then $f(z_1^0, z_2)$ is a continuous function of z_2 on the circumference $|z_2| = 1$;
- (b) if $|t_2| = 1$, then $f(z_1, t_2)$ is an analytic function of z_1 for $z_1 = th(t_2, \lambda)$ ($0 \leq t < 1$, $0 \leq \lambda \leq 2\pi$) and continuous on the boundary $z_1 = h(t_2, \lambda)$;
- (c) $f(h(t_2, \lambda), t_2)$ is continuous on the distinguished surface of \mathfrak{M} ; i. e.,

³ Hypothesis (c) can also be phrased in terms of conditions as to the boundedness of the first and second derivatives of $h(z_2^0, \lambda)$ with respect to λ , in a similar manner to that used in a recent paper by Bergmann and Marcinkiewicz, *Fundamenta Mathematicae*, vol. 33 (1939), pp. 75-94; in particular, Lemma 3, p. 80.

⁴ A statement of Hartogs' theorem will be found in Osgood, *Lehrbuch der Funktionentheorie*, vol. II, part 1, p. 199.

continuous in the variables λ and t_2 ($|t_2| = 1$). Then $f(z_1, z_2)$ can be continued analytically throughout \mathfrak{M} .

Proof. For each interior point of the product domain, the Cauchy integral formula for 1 c. v. applied to the variable z_2 gives

$$f(z_1, z_2) = \int_{|t_2|=1} \frac{f(z_1, t_2)}{t_2 - z_2} dt_2.$$

Moreover, for each $z_1 \in \mathfrak{R}^2$, the function $f(z_1, t_2)$ can by a second application of the Cauchy formula be written as

$$f(z_1, t_2) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{f(h(t_2, \lambda), t_2)}{h(t_2, \lambda) - z_1} \frac{\partial h(t_2, \lambda)}{\partial \lambda} d\lambda.$$

Combining these two one has for all points of the product domain the integral representation

$$(10) \quad f(z_1, z_2) = \frac{-1}{4\pi^2} \int_{|t_2|=1} \frac{dt_2}{t_2 - z_2} \int_0^{2\pi} \frac{f(h(t_2, \lambda), t_2)}{h(t_2, \lambda) - z_1} \frac{\partial h(t_2, \lambda)}{\partial \lambda} d\lambda.$$

But this last expression is the generalized Cauchy integral for the domain \mathfrak{M} , as given by Bergmann (B_3), and thus represents an a. f. 2 c. v. defined throughout \mathfrak{M} . Since the integral agrees with the original function in the product domain, it represents the analytic continuation of $f(z_1, z_2)$ over \mathfrak{M} .

5. Convergence in \mathfrak{M} -domains. Because the theory of two complex variables possesses no analogue to the Riemann mapping theorem it is not possible to pass directly from results for the bicylinder to statements concerning \mathfrak{M} -domains. An indirect method of surmounting this difficulty is to make use of a *domain of comparison* (B_4)—in this case, of a small bicylinder contained in the \mathfrak{M} -domain—and to show that certain hypotheses as to convergence on the distinguished surface of the \mathfrak{M} -domain imply conditions as to convergence on the bicylinder to which Theorem 3a is applicable. It is known by Lemma 1 that any f. 2 c. v., analytic in the bicylinder, which satisfies certain hypotheses on the distinguished surface of the \mathfrak{M} -domain can be continued analytically throughout the \mathfrak{M} -domain. The first step is to insure the existence of a bicylinder contained in \mathfrak{M} , by the introduction of suitable *normal coördinates* (B_4). Such a system of coördinates is given by the transformations

$$(11) \quad z_1^* = \frac{z_1 - h(z_2, \lambda_0)}{\left(\frac{\partial h(z_2, \lambda)}{\partial \lambda}\right)_{\lambda=\lambda_0}} + r, \quad z_2^* = z_2.$$

These take all points $z_1 = h(z_2, \lambda_0)$ into $z_1^* = r$ and all the inner normals to the curves $h(z_2^0, \lambda)$, taken at the point λ_0 , into the direction of the negative real axis. Thus by the use of these coördinates it may without loss of generality be assumed that for a particular value $\lambda = \text{const.}$, say $\lambda = 0$, the value of $h(z_2, 0)$ is independent of z_2 and that the inner normal to the curve $h(z_2^0, \lambda)$ ($z_2^0 = \text{const.}$) has at the point $\lambda = 0$ a direction independent of the value of z_2^0 . Since by hypothesis (c) on the two dimensional sections $z_2 = \text{const.}$ of the \mathfrak{M} -domain the boundaries of these sections all have radii of curvature greater than or equal to some positive number r , it may now be supposed that the \mathfrak{M} -domain contains a bicylinder $|z_1| \leq r$, $|z_2| \leq 1$ which is tangent to the boundary of \mathfrak{M} along the two dimensional surface $z_1 = h(z_2, 0)$.

As a further preliminary, it is necessary to state some results for the one variable case. Given a two dimensional simply-connected domain \mathfrak{G}^2 whose boundary g^1 satisfies the conditions imposed on the boundaries of the sections $z_2 = \text{const.}$ of the \mathfrak{M} -domain. Then it is possible, given a sufficiently small arc α^1 of g^1 , to describe a circle \mathfrak{R}^2 with center in \mathfrak{G}^2 whose circumference cuts g^1 only at the end-points of α^1 and which has a radius not less than the distance between these end-points; so that if b^1 denotes the arc of the circumference of \mathfrak{R}^2 which is subtended by α^1 and lies outside \mathfrak{G}^2 , then the central angle subtended by b^1 is not greater than $\pi/3$. By Carleman's extension principle (N, p. 63), the following inequalities are valid for all $z \in \mathfrak{G}^2 \cdot \mathfrak{R}^2$:

$$\omega(z, \alpha^1, \mathfrak{G}^2) > \omega(z, \alpha^1, \mathfrak{G}^2 \cdot \mathfrak{R}^2) \geq \omega(z, b^1, \mathfrak{R}^2);$$

so that the set $E[\omega(z, b^1, \mathfrak{R}^2) > \mu]$ is contained in the set $E[\omega(z, \alpha^1, \mathfrak{G}^2) > \mu]$. (Since there is little chance of confusion, ω is here used, as usual, to denote the harmonic measure specified by its argument). But the equipotential $\omega(z, b^1, \mathfrak{R}^2) = \mu$ is known (N, p. 7) to be the circular arc interior to \mathfrak{R}^2 whose end-points are the same as those of b^1 and which makes an angle $(1 - \mu)\pi$ with b^1 . In particular, by reason of the above hypotheses, a semi-circle whose end-points coincide with the common end-points of α^1 and b^1 makes an angle not greater than $2\pi/3$ with b^1 , so that for this case $\mu \geq 1/3$. Applying the one variable form of the two-constant theorem, we thus have the following result:

LEMMA 2. *Given $f(z)$ defined and regular in a domain \mathfrak{G}^2 whose boundary g^1 has at all points a positive radius of curvature and is such that about any sufficiently small arc α^1 of g^1 it is possible to describe a circle whose center is in \mathfrak{G}^2 , whose circumference cuts g^1 only at the end-points of α^1 , and which has a radius not less than the distance between these end-points. Then if $|f(z)| \leq 1$ on g^1 and $|f(z)| < \epsilon$ ($0 < \epsilon < 1$) on α^1 , one has $|f(z)| < \epsilon^{1/3}$*

at all points of the domain bounded by α^1 and the semi-circle whose end-points coincide with those of α^1 .

Using this last result, it now becomes possible to pass from a majorant on the distinguished surface of an \mathfrak{M} -domain to a majorant on the distinguished surface of the bicylinder to be used as a domain of comparison, and thus to use the result already obtained on convergence in the bicylinder to obtain a result on convergence in the \mathfrak{M} -domain. In what follows, the normal coördinates introduced in equations (11) with $\lambda = 0$ will be used without further explicit mention of the fact.

Still considering the one variable case, let g^1 be any curve $z_1 = h(z_2^0, \lambda)$ and let $z_1 = h(z_2^0, 0)$ be one end-point of the arc α^1 . Let a circle \mathfrak{C}^2 of fixed radius r , where r is the lower bound of the radii of curvature of g^1 (positive, by hypothesis), be drawn tangent to g^1 at $z_1 = h(z_2^0, 0)$ and lying in \mathfrak{G}^2 . Then the semi-circle \mathfrak{C}^2 whose end-points coincide with those of α^1 cuts off an arc on the circumference \mathfrak{C}^1 of \mathfrak{C}^2 whose length is certainly greater than half the length of α^1 , provided the distance d between the end-points of α^1 is not greater than r . For let c^1 be the arc cut off on \mathfrak{C}^1 , and let a and e be the length of α^1 and c^1 respectively. Obviously, the most unfavorable case is for $d = r$ and for α^1 coinciding with the tangent to g^1 at $z_1 = h(z_2^0, 0)$. In this case $d \leq a < \pi d/3$ and $e \geq \pi d/4$, so that $a \leq 4e/3$ and a fortiori $a < 2e$.

To summarize these results: By the hypotheses on the \mathfrak{M} -domain and by the use of the appropriate normal coördinates it is possible to assume without loss of generality that \mathfrak{M} contains a bicylinder $|z_1| \leq r$, $|z_2| \leq 1$ (r a fixed positive quantity) to which it is tangent along the two dimensional surface $z_1 = h(z_2, 0) = r$, $|z_2| \leq 1$. Moreover, to any arc $\alpha^1 = (h(z_2, \lambda), r)$ ($\lambda > 0$) on a section $z_2 = \text{const.}$, there corresponds an arc $(re^{i\theta}, r)$ ($\theta > 0$) cut off on $|z_1| = r$ by a semi-circle whose end-points coincide with those of α^1 , the length of this latter arc being greater than half the length of α^1 . Thus the set

$$(12) \quad \mathfrak{S}^2 = E \left[\frac{2a\delta [|h(z_2, \lambda) - r|^2 (1 + \text{Re} h(z_2, \lambda^*) - 2r(\text{Im} h(z_2, \lambda))(\text{Im} h(z_2, \lambda^*)))]}{|1 + h(z_2, \lambda^*)|^2 |h(z_2, \lambda) - r|^2} \right. \\ \left. + \frac{4y_2^2}{|1 - z_2|^4} < 0, \right. \\ \left. |z_2| = 1, \lambda^* = \text{const.}, \lambda^* \geq \lambda \geq 0, |h(z_2, \lambda^*) - r| > 2r |e^{i\theta} - 1| \right]$$

corresponds in this manner to a set on the distinguished surface of the bicylinder which includes the set (7) of Theorem 3a. Hence, using the results of Theorem 3a, it is now possible to state the following result for the case of an \mathfrak{M} -domain:

THEOREM 4. *Given $f(z_1, z_2)$ defined and regular in an \mathfrak{M} -domain given by equation (9) and satisfying the hypotheses (a), (b), and (c) following equation (9). If $|f(z_1, z_2)| \leq 1$ on the distinguished surface $z_1 = h(z_2, \lambda)$, $|z_2| = 1$ of \mathfrak{M} and $|f(z_1, z_2)| < \epsilon$ for $\{z_1, z_2\}$ belonging to the set \mathfrak{Z}^2 given by (12), then $|f(z_1, z_2)| < \epsilon^{u/3}$ for $\{z_1, z_2\}$ belonging to the set (8) of Theorem 3a.*

This theorem has an interpretation for convergence wholly analogous to that previously given for Theorem 3.

6. Extensions and applications. These results can be extended in at least two directions. First, the hypotheses on the set \mathfrak{Z}^2 of Theorem 4 can undoubtedly be made sharper if necessary, and the hypothesis that $f(z_1, z_2)$ be regular in the entire \mathfrak{M} -domain can be lightened in accordance with Lemma 1 to the supposition that the function is merely defined, bounded, and continuous on the distinguished surface of M and regular in the intersection of \mathfrak{M} and a product domain which includes the particular two dimensional surface $z_1 = h(z_2, 0)$, $|z_2| \leq 1$ where convergence is to be studied. Second, reverting to Theorem 3, it is possible to use n overlapping parabolas and to suppose, for example, that $|f(z_1, z_2)| \leq 1$ on the real plane $y_1 = y_2 = 0$ and $|f(z_1, z_2)| < n\epsilon$ in the overlapping portions of the parabolas. The doubly harmonic measure of the resulting region is obtained merely by addition of the doubly harmonic measures of the individual parabolas.

One of the interesting applications of the theory of functions of two complex variables is to the theory of pseudo-conformal mapping; i. e., mapping of a four dimensional domain by a pair of analytic functions of two complex variables. The results here stated apply only to a single function, but can easily be applied, in connection with some results of Bergmann (B_5), to the study of convergence to the boundary of the pseudo-conformal map of a domain. I hope to discuss these matters further at a later time.

UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA.

ON THE NON-EXISTENCE OF THE EUCLIDEAN ALGORITHM IN CERTAIN QUADRATIC NUMBER FIELDS.*

By ALFRED BRAUER.

Introduction. Let P be the field of rational numbers, and m be a rational integer which is not divisible by the square of a prime. If for any pair of integers α, β with $\beta \neq 0$ of $P(m^{1/2})$ a third integer γ of this field can be determined such that

$$(1) \quad |N(\alpha - \beta\gamma)| < |N(\beta)|,$$

where $N(\alpha)$ is the norm of α , we say¹ that the Euclidean algorithm exists in the field $P(m^{1/2})$ or that the field is Euclidean.

The problem of determining all Euclidean quadratic fields has not been solved completely, although this question has been studied a great deal during the last few years. In this paper I prove that the algorithm does not exist in certain cases in which the question has remained unsolved till now.

If a field is Euclidean, then the greatest common divisor exists for any pairs of integers of this field; thus it is necessary that the class number is equal to 1. Dedekind² remarked that this condition is not sufficient, because the class number is equal to 1 in the field $P(\sqrt{-19})$, although this field is not Euclidean.

For imaginary quadratic fields it was shown by L. E. Dickson³ that the Euclidean algorithm exists only in the cases $m = -1, -2, -3, -7$, and -11 . For real quadratic fields the question has not yet been solved completely. For

$$(2) \quad m = 2, 3, 5, 6, 7, 11, 13, 17, 19, 21, 29, 33, 37, 41, 57,$$

the algorithm exists. This follows from the investigations of O. Perron,⁴ A. Oppenheim,⁵ R. Remak,⁶ E. Berg,⁷ and N. Hofreiter.⁸ I. Schur⁹ remarked

* Received October 13, 1939.

¹ Cf. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford (1938), pp. 212-217.

² P. G. Lejeune Dirichlet, *Vorlesungen über Zahlentheorie*, herausgegeben von R. Dedekind, 4. Aufl. Braunschweig (1894), p. 451.

³ *Algebren und ihre Zahlentheorie*, Zürich u. Leipzig (1927), pp. 150-151.

⁴ "Quadratische Zahlkörper mit Euklidischem Algorithmus," *Mathematische Annalen*, vol. 107 (1932), pp. 489-495.

⁵ "Quadratic fields with and without Euclid's algorithm," *Mathematische Annalen*, vol. 109 (1934), pp. 349-352.

⁶ "Über den Euklidischen Algorithmus in reell-quadratischen Zahlkörpern," *Jahresbericht d. Deutschen Mathematiker-Vereinigung*, vol. 44 (1934), pp. 238-250.

that the algorithm does not exist for $m = 47$. A. Oppenheim¹⁰ proved the non-existence for $m = 23$ and $m = 53$, N. Hofreiter¹¹ for $m \equiv 14 \pmod{24}$, and also¹² for $m = 77$ and for $m \equiv 21 \pmod{24}$ with $m > 21$, E. Berg¹³ and J. Fox Keston^{13a} for $m \not\equiv 1 \pmod{4}$ except in the cases (2). Some of these results are also proved in Hardy and Wright's book mentioned above in the footnote 1. H. Behrbohm and L. Rédei¹⁴ showed that, excepting the cases (2), the algorithm can only exist in the following three cases (p and q denote primes)

$$\text{I. } m = p \equiv 13 \pmod{24},$$

$$\text{II. } m = p \equiv 1 \pmod{8},$$

$$\text{III. } m = pq \text{ with } p \equiv q \equiv 3 \pmod{8} \text{ or } p \equiv q \equiv 7 \pmod{8}.$$

Using analytical methods, P. Erdős and Ch. Ko¹⁵ proved that the algorithm does not exist in the cases I and II, if m is sufficiently large. The corresponding fact in the case III was shown by H. Heilbronn.¹⁶ Finally, L. Schuster¹⁷ proved that in the case III, the algorithm exists at most for $m \equiv 1 \pmod{24}$ except for $m = 33$ and $m = 57$.

In this paper I improve the theorem of Erdős and Ko for the case I. I show by elementary methods that here the algorithm cannot exist for $p > 109$. In the cases $p = 13$ and $p = 37$ the algorithm exists; whether or not the fields $P(\sqrt{61})$ and $P(\sqrt{109})$ are Euclidean, I cannot decide.

In their paper mentioned above, Erdős and Ko prove that the algorithm does not exist in the case I, if the two least quadratic non-residues u and v which are odd primes satisfy the condition

⁷ "Über die Existenz eines Euklidischen Algorithmus in quadratischen Zahlkörpern," *Kungl. Fysiografiska Sällskapets i Lund Föreläsningar*, vol. 5 (1935), Nr. 5.

⁸ "Quadratische Körper mit und ohne Euklidischen Algorithmus," *Monatshefte für Mathematik und Physik*, vol. 42 (1935), pp. 397-400.

⁹ Cf. loc. cit. 5), p. 351.

¹⁰ Loc. cit. 5).

¹¹ "Quadratische Zahlkörper ohne Euklidischen Algorithmus," *Mathematische Annalen*, vol. 110 (1935), pp. 195-196.

¹² Loc. cit. 8).

¹³ Loc. cit. 7).

^{13a} "Existence of a Euclidean algorithm in quadratic fields," *Thesis Yale University* (1935); cf. *Bulletin of the American Mathematical Society*, vol. 41 (1935), p. 186.

¹⁴ "Der Euklidische Algorithmus in quadratischen Zahlkörpern," *Journal f. d. reine u. angewandte Mathematik*, vol. 174 (1936), pp. 192-205.

¹⁵ "Note on the Euclidean algorithm," *Journal of the London Mathematical Society*, vol. 13 (1938), pp. 3-8.

¹⁶ "On Euclid's algorithm in real quadratic fields," *Proceedings of the Cambridge Philosophical Society*, vol. 34 (1938), pp. 521-526.

¹⁷ "Reellquadratische Zahlkörper ohne Euklidischen Algorithmus," *Monatshefte f. Mathematik u. Physik*, vol. 47 (1938), pp. 117-127.

$$(3) \quad 3uv < p.$$

Then they use analytical methods for proving that (3) is satisfied for all primes which are sufficiently large. I here prove (3) for all primes $p > 421$ of the form $24n + 13$ in an elementary manner, using the inequality for the least odd quadratic non-residue u modulo a prime p of the form $8n \pm 5$,

$$(4) \quad u < 2\{(4p)^{2/5} + (4p)^{1/5}\} + 1,$$

which I had obtained by elementary methods in a former paper.¹⁸

After dealing briefly with the method of Erdős and Ko in § 1, I prove in § 2 that the algorithm does not exist in $P(p^{1/2})$, if p is of the form $24n + 13$ and $v > 8u$. If, however, $v < 8u$, then the non-existence of the algorithm for sufficiently large p follows immediately from (4). The limit for p for which this holds can easily be given. In order to replace it by a smaller number, I show in § 3 that the Euclidean algorithm does not exist, if $p = 24n + 13 > 12696$ and $v > 6u$. From these theorems, the non-existence of the algorithm follows for the primes of this type, if $24u^2 < p$, or if $18u^2 < p$ and $p > 12696$.

In § 4, I improve (4) to

$$\begin{aligned} u &< 2^{3/5}p^{2/5} + 2^{-(6/5)} \cdot 25p^{1/5} + 3 \quad \text{for } p = 8n + 5, \\ u &< 2p^{2/5} + 4\frac{1}{2}p^{1/5} + 7 \quad \text{for } p = 8n + 3. \end{aligned}$$

For $p = 8n + 5$ this is still further improved. On the basis of these theorems we obtain the result in § 5 that the algorithm does not exist in $P(p^{1/2})$ for $p = 24n + 13 > 3\,300\,000$. The primes below this limit must be treated directly. Using the above theorems we can see that the algorithm does not exist for $p > 109$. This direct treatment requires long computation, even if properly arranged. In these computations I have been assisted by my mother and my wife.

1. The method of Erdős and Ko. Erdős and Ko prove the following theorems:

THEOREM 1. *For a prime p of the form $4n + 1$, the Euclidean algorithm cannot exist in $P(p^{1/2})$, if p can be written in the form*

$$(5) \quad p = q_1m_1 + q_2m_2,$$

where m_1, m_2, q_1, q_2 are all positive and quadratic non-residues (mod p), and where the q_i are odd primes which divide q_im_i to an odd power for $i = 1, 2$.

Proof. We write the condition (1) in the form

¹⁸ "Über den kleinsten quadratischen Nichtrest," *Mathematische Zeitschrift*, vol. 33 (1930), pp. 161-176.

$$(6) \quad |N(\alpha/\beta - \gamma)| < 1.$$

Suppose now

$$(7) \quad \begin{cases} \alpha/\beta = r + s p^{1/2}, \\ \gamma = \frac{1}{2}(x + y p^{1/2}), \end{cases}$$

where r, s are rational and x, y rational integers with $x \equiv y \pmod{2}$. This is possible, since γ is any integer of the field $\mathbf{P}(p^{1/2})$ and $p \equiv 1 \pmod{4}$. From (6) and (7) we obtain

$$(8) \quad |(x - 2r)^2 - p(y - 2s)^2| < 4.$$

Consequently, if for a pair of rational numbers r and s it is impossible to determine the rational integers x and y such that the condition (8) is satisfied, the field is not Euclidean. Since $q_1 m_1$ is a quadratic residue, the congruence $z^2 \equiv q_1 m_1 \pmod{p}$ has a solution z_1 .

We now choose

$$r = 0, \quad s = z_1/p,$$

and it follows from (8) that

$$|px^2 - (py - 2z_1)^2| < 4p.$$

Since here the left-hand side is congruent to $4z_1^2 \pmod{4p}$, we have either

$$(9) \quad px^2 - (py - 2z_1)^2 = -4q_1 m_1,$$

or, by (5),

$$(10) \quad px^2 - (py - 2z_1)^2 = 4p - 4q_1 m_1 = 4q_2 m_2.$$

We have to show that (9) is not possible. Suppose first that $x \equiv 0 \pmod{q_1}$. Then also $py - 2z_1 \equiv 0 \pmod{q_1}$ and q_1 divides the left-hand side of (9) to an even power, but the right-hand side to an odd power. This is impossible. Suppose now $x \not\equiv 0 \pmod{q_1}$. Then it follows from (9) that px^2 is a quadratic residue of q_1 . This is impossible, because

$$(p/q_1) = (q_1/p) = -1.$$

Thus (9) is impossible. In the same way it follows that (10) cannot be solvable.

THEOREM 2. *Let p be a prime of the form $24n + 13$. If u and v are the two least quadratic non-residues \pmod{p} , which are odd primes, and if*

$$(11) \quad 3uv < p,$$

then the Euclidean algorithm does not exist in $\mathbf{P}(p^{1/2})$.

Proof. If we set

$$(12) \quad p = 3uv + 2b_1$$

and

$$(13) \quad p = uv + 2b_2,$$

then b_1 and b_2 are positive integers, because of (11). For primes $p = 24n + 13$, the number 2 is a quadratic non-residue and 3 a quadratic residue. Hence, from (12) it follows that

$$(2b_1/p) = (-3uv/p) = (uv/p) = 1.$$

Consequently, b_1 is a non-residue. In an analogous manner it follows from (13) that b_2 is a non-residue. Further, from (12) and (13), we have

$$p = 3b_2 - b_1.$$

Therefore, one of the two numbers b_1, b_2 is odd. Let us first assume that b_1 is odd. Since b_1 was a quadratic non-residue, there exists at least one odd prime q which is a non-residue (mod p) and which divides b_1 to an odd power. Because of (12), the number q is different from u and v . We set in Theorem 1

$$q_1 = u, \quad m_1 = 3v, \quad q_2 = q, \quad m_2 = 2b_1/q.$$

Then (12) yields a representation of p which satisfies the conditions of Theorem 1. Therefore the algorithm does not exist in $P(p^{1/2})$ in this case.

If, however, b_2 is odd, then there exists at least one odd prime q' which is a quadratic non-residue (mod p) and which divides b_2 to an odd power. From Theorem 1 for

$$q_1 = u, \quad m_1 = v, \quad q_2 = q', \quad m_2 = 2b_2/q',$$

and from (13) it then follows that the algorithm does not exist in $P(p^{1/2})$ in this case.

The Theorems 1 and 2 will be used in the following. The proofs are given here again, in the first place in order to show that they are elementary, in the second because the Theorem 2 is not given explicitly in the paper of Erdős and Ko. There it is assumed instead of (11) that the three least quadratic non-residues u, v, w (mod p) which are odd primes satisfy the condition

$$uvw < p^{1-\eta}$$

where $\eta < .001$ is a positive constant. But for the primes of the form $24n + 13$, Erdős and Ko actually use only the weaker assumption (11). In my paper it will be important that the condition (11) is sufficient in this case.

2. Elementary proof of the theorem for large p . We first prove the following theorem:

THEOREM 3. *Let p be a prime of the form $24n + 13$. If the two least quadratic non-residues u, v (mod p), which are odd primes, satisfy the condition*

$$(14) \quad v > 8u,$$

then there does not exist an Euclidean algorithm in $P(p^{1/2})$.

Proof. We have nothing to prove for $p = 13, 37$, and 61 , since (14) is not true for these primes. We assume therefore $p \geq 109$.

The least odd quadratic non-residue u modulo a prime p of the form $8n + 5$ satisfies the condition

$$(15) \quad u < (p + 4)^{\frac{1}{2}} + 2,$$

as I have shown in the paper mentioned above.¹⁰ Hence, because of $p > 96$

$$(16) \quad p - 8u > p - 8(p + 4)^{\frac{1}{2}} - 16 > 0,$$

since

$$(p - 16)^2 = p^2 - 32p + 256 > 64p + 256.$$

Let $2ku$ be the largest multiple of $2u$ which is less than p . The following eight even numbers

$$(17) \quad \begin{aligned} &2(k-3)u, 2(k-2)u, 2(k-1)u, 2ku, 2(k+1)u, \\ &2(k+2)u, 2(k+3)u, 2(k+4)u \end{aligned}$$

lie in the interval $\{p - 8u \cdots p + 8u\}$ and are therefore positive because of (16). They form an arithmetical progression with the difference $2u$. It follows that exactly two of these numbers are divisible by 4 and not by 8; the difference of these two numbers is $8u$. This implies that at least one of them, say $4lu$, is not divisible by u^2 . Then

$$(18) \quad (l, 2u) = 1.$$

Furthermore, since $4lu$ is one of the numbers (17), we have

$$(19) \quad p - 8u < 4lu < p + 8u,$$

$$(20) \quad |p - 4lu| < 8u.$$

On the other hand, $|p - 4lu|$ is an odd integer less than $8u$ because of (20), hence less than v because of (14). It follows that $|p - 4lu|$ is a quadratic residue (mod p), since $|p - 4lu|$ is not divisible by u , and all the odd positive integers less than v , which are not divisible by u , are quadratic residues. Then lu also is a quadratic residue, and therefore l a quadratic non-residue (mod p). Because of (18), l contains at least one odd prime w which is different from u and a quadratic non-residue (mod p). Consequently, $v \leq w$, hence because of (19)

$$4uv \leq 4uw \leq 4lu < p + 8u,$$

$$3uv + uv < p + 8u,$$

$$3uv + u(v - 8) < p.$$

Thus, because of (14)

$$3uv < p.$$

¹⁰ *Loc. cit.* 18), Satz 2.

Theorem 2 now shows that the Euclidean algorithm cannot exist in $P(p^{1/2})$. This proves Theorem 3.

If, on the other hand, the assumption (14) is not satisfied, i. e., if $v < 8u$, then (4) implies the inequality

$$(21) \quad 3uv < 24u^2 < 24\{2(4p)^{2/5} + 2(4p)^{1/5} + 1\}^2.$$

However, if p is sufficiently large, we have

$$(22) \quad 24\{2(4p)^{2/5} + 2(4p)^{1/5} + 1\}^2 < p.$$

For all values of p for which (22) holds we have because of (21)

$$3uv < p.$$

According to Theorem 2 the Euclidean algorithm does not exist in $P(p^{1/2})$ for these p in the case $v < 8u$ we are considering. In connection with Theorem 3 this yields the theorem of Erdős and Ko.

THEOREM 4. *If p is a sufficiently large prime of the form $24n + 13$, then there does not exist an Euclidean algorithm in $P(p^{1/2})$.*

Since (4) had been obtained by elementary methods, we have given a proof which is free of analytical methods. More exactly, we see that the algorithm does not exist, when (22) is satisfied. We may easily obtain a lower bound for p from which (22) holds. We do not give the computation, since we shall later obtain a still smaller value of this lower bound.

As an immediate consequence of Theorem 3 we have

THEOREM 5. *If the least odd quadratic non-residue u modulu a prime p of the form $24n + 13$ satisfies the condition $24u^2 < p$, then there does not exist an Euclidean algorithm in $P(p^{1/2})$.*

Proof. Let again u and v be the least quadratic non-residues (mod p) which are odd primes, $u < v$. If $v > 8u$, the statement follows from Theorem 3. If, however, $v < 8u$, then

$$3uv < 24u^2 < p$$

and the theorem follows from Theorem 2.

3. Improvement of Theorem 3. The Theorem 3 can be improved in the following manner:

THEOREM 6. *If $p > 12696$ is a prime of the form $24n + 13$, and if the two least quadratic non-residues u and v (mod p) which are odd primes satisfy the condition*

$$(23) \quad v > 6u,$$

then there does not exist an Euclidean algorithm in $P(p^{1/2})$.

Proof. If $u \leq 23$, then $24u^2 \leq 12696$, and the statement follows from Theorem 5. We assume therefore that

$$(24) \quad u \geq 29.$$

As in the proof of Theorem 3, let $2ku$ be the greatest integral multiple of $2u$ less than p . We take here the following four integers

$$(25) \quad 2(k-1)u, \quad 2ku, \quad 2(k+1)u, \quad 2(k+2)u$$

which belong to (17). These integers lie in the interval $\{p-4u \cdots p+4u\}$. They are all positive because of (16), since $p > 12696 > 96$. There is exactly one of them which is divisible by 4, but not by 8. Suppose that this is the number $4lu$. Analogously to (19) and (20), we obtain from (25)

$$(26) \quad p-4u < 4lu < p+4u,$$

$$(27) \quad |p-4lu| < 4u < 6u.$$

Further, l is odd. If we have

$$(28) \quad (l, u) = 1,$$

in accordance with (18), then the statement follows from (23), (26), (27), and (28) in complete analogy with the proof of Theorem 3.

On the other hand, let us suppose that

$$(l, u) > 1.$$

Then we have

$$(l, u) = u,$$

since u is a prime; hence

$$(29) \quad 4lu \equiv 0 \pmod{u^2}.$$

From (15) it follows that

$$(30) \quad p-24u > p-24(p+4)^{\frac{1}{2}}-48 > 0$$

since $p > 672$, and therefore

$$(p-48)^2 = p^2 - 96p + 2304 > 576p + 2304.$$

We consider the interval

$$(31) \quad I = \{p-24u \cdots p\}.$$

There lie exactly 4 odd multiples of $3u$ in I . Let

$$(32) \quad 3su, \quad 3(s+2)u, \quad 3(s+4)u, \quad 3(s+6)u$$

be these multiples. The even integers

$$(33) \quad p-3(s+6)u, \quad p-3(s+4)u, \quad p-3(s+2)u, \quad p-3su$$

form an arithmetical progression with the difference $6u$. It follows that

exactly one of the numbers (33) is divisible by 4 and not by 8. Suppose that this is $p - 3tu$; then we have

$$(34) \quad \frac{1}{4}(p - 3tu) \equiv 1 \pmod{2}.$$

The integer $3tu$ belongs to the numbers (32), hence we have

$$(35) \quad t \equiv 1 \pmod{2}$$

and

$$(36) \quad 0 < p - 24u < 3tu < p$$

according to (31) and (30).

Moreover $3tu$ and $4lu$ both lie in the interval $\{p - 24u \cdots p + 4u\}$ because of (26) and (36). Their difference then is at most equal to $28u$. Because of (24), we have

$$(37) \quad |3tu - 4lu| \leq 28u < u^2.$$

According to (35), $3tu$ is odd, and therefore different from $4lu$. It follows from (37) that $3tu$ and $4lu$ are not both divisible by u^2 . Hence we obtain

$$(38) \quad 3tu \not\equiv 0 \pmod{u^2}$$

by (29). Furthermore, we have

$$(39) \quad 0 < \frac{1}{4}(p - 3tu) < 6u < v$$

because of (36) and (23).

The integer $\frac{1}{4}(p - 3tu)$ is odd, not divisible by u , and less than v because of (34) and (39). Consequently, it is a quadratic residue \pmod{p} ; so is $3tu$. It follows that $3t$ and t are quadratic non-residues. But t was odd because of (35), positive because of (36), and not divisible by u , according to (38). Then t contains at least one odd prime factor w which is a quadratic non-residue \pmod{p} but different from u . For it, we have $v \leq w$, and therefore, because of (36)

$$3uv \leq 3uw \leq 3ut < p.$$

The statement of Theorem 6 follows now from Theorem 2.

From Theorem 6 we obtain at once the following theorem which improves Theorem 4:

THEOREM 7. *If $p > 12696$ is a prime of the form $24n + 13$ and if the least quadratic non-residue $u \pmod{p}$, which is an odd prime, satisfies the condition $18u^2 < p$, then there does not exist an Euclidean algorithm in $\mathbf{P}(p^{1/2})$.*

4. Estimates for the least odd quadratic non-residue. In my paper mentioned in the introduction, I have shown that the least odd quadratic non-residue u for a prime of the form $8n \pm 5$ satisfies the inequality

$$(40) \quad u < 2\{(4p)^{2/5} + (4p)^{1/5}\} + 1.$$

It was mentioned there that this relation can still be improved for primes of the form $8n + 5$. In this manner, we may obtain

$$(41) \quad u < 2\{(2p)^{2/5} + (2p)^{1/5}\} + 1.$$

We now have to improve (40) and (41) still further.

THEOREM 8. *The least odd quadratic non-residue u modulo a prime p satisfies*

$$(42) \quad \begin{cases} u < 2^{3/5}p^{2/5} + 2^{-(6/5)} \cdot 25p^{1/5} + 3 & \text{for } p = 8n + 5, \\ u < 2p^{2/5} + 4^{9/2}p^{1/5} + 7 & \text{for } p = 8n + 3. \end{cases}$$

Proof. We have nothing to prove for $u = 3$ and $u = 5$; hence we assume $u \geq 7$. The even numbers

$$p + 1, p + 3, \dots, p + u - 2$$

are quadratic residues. For a p of the form $8n + 5$, the numbers

$$p - 1, p - 3, \dots, p - u + 2$$

are also quadratic residues. Let U denote the interval $\{p \cdots p + u - 1\}$ if p is of the form $8n + 3$, and the interval $\{p - u + 1 \cdots p + u - 1\}$ for p of the form $8n + 5$. Then all even integers of U are quadratic residues. If z is an arbitrary odd integer such that

$$(43) \quad \begin{cases} 1 \leq z < u/2 & \text{for } p = 8n + 3, \\ 1 < z < u/2 & \text{for } p = 8n + 5, \end{cases}$$

then U contains integral multiples of $2z$. Let

$$(44) \quad k \cdot 2z, (k + 1)2z, \dots, (k + l - 1)2z$$

be those multiples of $2z$; then

$$(45) \quad k \leq \left[\frac{p}{2z} \right] + 1.$$

All the numbers (44) are quadratic residues as even numbers of U , 2 is a non-residue, and z a quadratic residue because of (43). This implies that the numbers

$$(46) \quad k, k + 1, \dots, k + l - 1$$

form a sequence of l non-residues. None of them is therefore a square. Hence we may find a positive integer a such that

$$(47) \quad a^2 < k \leq k + l - 1 < (a + 1)^2.$$

Then it follows from (45) that

$$(48) \quad \begin{aligned} a^2 &\leq \left[\frac{p}{2z} \right], \\ a &< \left(\frac{p}{2z} \right)^{\frac{1}{2}}. \end{aligned}$$

We divide now the interval $A = \{a^2 \cdots (a+1)^2\}$ into parts; we have to distinguish between two cases.

I. a even:

We determine the positive integer l' such that

$$(49) \quad (a+1)^2 - (2l')^2 > a^2 > (a+1)^2 - (2l'+2)^2.$$

This can always be done since an even square number cannot equal the difference of an odd and an even square number and since $a \geq 2$. Then

$$(50) \quad \begin{aligned} (2l')^2 &< 2a+1, \\ 2l' &\leq [\sqrt{2a}]. \end{aligned}$$

The points

$$(51) \quad (a+1)^2 - (2v)^2 \quad (v=1, 2, \dots, l')$$

divide the interval A into subintervals. The distance between two such consecutive points is

$$(52) \quad (a+1)^2 - (2v)^2 - [(a+1)^2 - (2v+2)^2] = 8v+4.$$

The largest of the subintervals of A is therefore either the interval

$$I_{l'} = \{(a+1)^2 - (2l'-2)^2 \cdots (a+1)^2 - (2l')^2\},$$

or the interval

$$I_{l'+1} = \{(a+1)^2 - (2l')^2 \cdots a^2\}.$$

Because of (52), we find for the length $|I_{l'}|$ of $I_{l'}$

$$(53) \quad |I_{l'}| = 8l' - 4.$$

Since a was even, we have

$$a^2 - [(a+1)^2 - (2l'+2)^2] \equiv 3 \pmod{4}.$$

Because of (49), we find then

$$a^2 - [(a+1)^2 - (2l'+2)^2] \geq 3.$$

For the length $|I_{l'+1}|$ of $I_{l'+1}$ we have therefore

$$(54) \quad |I_{l'+1}| \leq 8l' + 4 - 3 = 8l' + 1$$

because of (52). If s' now denotes the maximal length of a subinterval of A , we obtain from (53), (54), and (50)

$$(55) \quad s' \leq 8l' + 1 \leq 4[\sqrt{2a}] + 1.$$

II. a odd:

In this case we determine an integer $l'' \geq 0$ such that

$$(56) \quad (a+1)^2 - (2l''+1)^2 > a^2 > (a+1)^2 - (2l''+3)^2.$$

This again is possible since the sum of two odd squares can not be a square. Then

$$(57) \quad \begin{aligned} 2l''+1 &< \sqrt{2a+1}, \\ 2l''+1 &\leq [\sqrt{2a}]. \end{aligned}$$

The points

$$(58) \quad (a+1)^2 - (2\nu+1)^2 \quad (\nu = 0, 1, 2, \dots, l'')$$

again divide the interval A into parts, and the distance of two consecutive points (58) is given by

$$(59) \quad (a+1)^2 - (2\nu+1)^2 - ((a+1)^2 - (2\nu+3)^2) = 8\nu + 8.$$

Consequently, if $l'' > 0$ then either the interval

$$I_{l''} = \{(a+1)^2 - (2l''-1)^2 \cdots (a+1)^2 - (2l''+1)^2\}$$

or the interval

$$I_{l''+1} = \{(a+1)^2 - (2l''+1)^2 \cdots a^2\}$$

is the largest of the subintervals of A . If $l'' = 0$ then $I_{l''+1}$ is the largest subinterval of A . If again $|I_{l''}|$ and $|I_{l''+1}|$ denote the lengths of $I_{l''}$ and $I_{l''+1}$, then because of (59)

$$(60) \quad |I_{l''}| = 8l''.$$

Since a was odd, we have

$$a^2 - ((a+1)^2 - (2l''+3)^2) \equiv 2 \pmod{4};$$

hence because of (56)

$$a^2 - ((a+1)^2 - (2l''+3)^2) \geq 2$$

and because of (59)

$$(61) \quad |I_{l''+1}| \leq 8l'' + 8 - 2 = 8l'' + 6.$$

Let s'' denote the maximal length of the subintervals of A . From (60), (61), and (57), it follows that

$$(62) \quad s'' \leq 8l'' + 6 \leq 4[\sqrt{2a}] + 2.$$

We set now

$$(63) \quad s = \text{Max}(s', s'').$$

In both cases I and II, the length of each subinterval of A is at most equal to s and we have, because of (55) and (62)

$$(64) \quad s \leq 4[\sqrt{2a}] + 2.$$

If we had now

$$(65) \quad u \geq \text{Max} \{a + 2 + [\sqrt{2a}], \epsilon zs\},$$

where

$$(66) \quad \begin{cases} \epsilon = 1 & \text{for } p = 8n + 5, \\ \epsilon = 2 & \text{for } p = 8n + 3, \end{cases}$$

then all the odd integers $\leq a + 1 + [\sqrt{2a}]$ would be quadratic residues. In the case that a is even, the odd integers

$$(67) \quad (a + 1)^2 - (2v)^2 = (a + 1 + 2v)(a + 1 - 2v)$$

for $v = 1, 2, \dots, t'$ would all be quadratic residues because of (50). Similarly, for odd a , the odd numbers

$$(68) \quad (a + 1)^2 - (2v + 1)^2 = (a + 1 + 2v + 1)(a + 1 - 2v - 1)$$

for $v = 0, 1, \dots, t''$ would all be quadratic residues because of (57).

In both cases, we have divided the interval A by the points (51) and (58) respectively which are quadratic residues according to (67) and (68). The length of each subinterval of A was smaller than or equal to s because of (63). It follows that each interval of length s which lies in A contains a quadratic residue, and hence A can not contain a sequence of s non-residues.

On the other hand, we have because of (65)

$$(69) \quad u \geq \epsilon zs.$$

In the case of a prime p of the form $8n + 3$, the interval U contained u consecutive integers, and then, according to (69) and (66) at least s complete residue systems mod $2z$ and hence at least s multiples of $2z$. For $p = 8n + 5$, there appear at least s multiples of z among the u numbers $p, p + 1, \dots, p + u - 1$ because of (69) and (66). The same is true for the u numbers $p, p - 1, \dots, p - u + 1$. But here, we have $z > 1$, according to (43), and this shows that z does not divide p . Thus, we have also in $U = \{p - u + 1, \dots, p + u - 1\}$ at least $2s$ consecutive multiples of z , hence s multiples of $2z$.

On the other hand, the number of multiples of $2z$ in U was equal to l because of (44), hence

$$l \geq s.$$

It follows from (46) and (47) that the interval A contains a sequence $k, k + 1, \dots, k + l - 1$ of at least s non-residues. This gives a contradiction which shows that (65) is not true. Hence

$$u < \text{Max} \{a + 2 + [\sqrt{2a}], \epsilon zs\}$$

and then according to (64)

$$u < \text{Max} \{a + 2 + [\sqrt{2a}], \epsilon z(4[\sqrt{2a}] + 2)\},$$

$$u \leq \text{Max} \{a + 1 + [\sqrt{2a}], \epsilon z(4[\sqrt{2a}] + 2) - 1\}.$$

Because of (48), we have then

$$(70) \quad u < \text{Max} \left\{ \left(\frac{p}{2z} \right)^{1/2} + 2^{1/2} \left(\frac{p}{2z} \right)^{1/4} + 1, \epsilon z \left(4 \cdot 2^{1/2} \left(\frac{p}{2z} \right)^{1/4} + 2 \right) - 1 \right\},$$

$$u < \text{Max} \left\{ \left(\frac{p}{2z} \right)^{1/2} + \left(\frac{2p}{z} \right)^{1/4} + 1, 4\epsilon(2pz^3)^{1/4} + 2\epsilon z - 1 \right\}.$$

For p of the form $8n + 5$, we assume first that $p > 2048$. For p of the form $8n + 3$, no restriction is necessary.

If we assume that

$$(71) \quad u \geq 2^{3/5} \epsilon^{2/5} p^{2/5} + 2^{9/5} \epsilon^{6/5} p^{1/5} (3 + 2^{-3} \epsilon^{-1}) + 4\epsilon - 1$$

and if we determine the odd integer z_0 so that

$$(72) \quad \frac{1}{4\epsilon} \left(\frac{p\epsilon}{2} \right)^{1/5} + 2 > z_0 > \frac{1}{4\epsilon} \left(\frac{p\epsilon}{2} \right)^{1/5},$$

then it follows from (71) and (72) that $\frac{1}{2}u > z_0$. For $p = 8n + 5$, we have $z_0 > 1$ because of (66), since $p > 2048$. For $z = z_0$, the conditions (43) are satisfied. Then it follows from (70) and (72) that

$$(73) \quad u < \text{Max} \left\{ \sqrt{2\epsilon p \left(\frac{p\epsilon}{2} \right)^{-(1/5)}} + \sqrt[4]{8p\epsilon \left(\frac{p\epsilon}{2} \right)^{-(1/5)}} + 1, \right.$$

$$4\epsilon \sqrt[4]{2p \left(\frac{1}{4\epsilon} \left(\frac{p\epsilon}{2} \right)^{1/5} + 2 \right)^3} + \frac{1}{2} \left(\frac{p\epsilon}{2} \right)^{1/5} + 4\epsilon - 1 \left. \right\},$$

$$u < \text{Max} \left\{ 2^{3/5} \epsilon^{2/5} p^{2/5} + 2^{4/5} \epsilon^{1/5} p^{1/5} + 1, 4\epsilon p^{1/4} \left(\frac{1}{32\epsilon^3} \left(\frac{p\epsilon}{2} \right)^{3/5} \right. \right.$$

$$\left. \left. + \frac{3}{4\epsilon^2} \left(\frac{p\epsilon}{2} \right)^{2/5} + \frac{6}{\epsilon} \left(\frac{p\epsilon}{2} \right)^{1/5} + 16 \right)^{1/4} + \frac{1}{2} \left(\frac{p\epsilon}{2} \right)^{1/5} + 4\epsilon - 1 \right\}.$$

We state now that

$$(74) \quad \frac{1}{32\epsilon^3} \left(\frac{p\epsilon}{2} \right)^{3/5} + \frac{3}{4\epsilon^2} \left(\frac{p\epsilon}{2} \right)^{2/5} + \frac{6}{\epsilon} \left(\frac{p\epsilon}{2} \right)^{1/5} + 16 < 2^{-(28/5)} \epsilon^{-(12/5)} p^{3/5}$$

$$+ 3 \cdot 2^{-(12/5)} \epsilon^{-(8/5)} p^{2/5} + 27 \cdot 2^{-(11/5)} \epsilon^{-(4/5)} p^{1/5} + 27 + 81 \cdot 2^{-(4/5)} \epsilon^{4/5} p^{-(1/5)}$$

$$= (2^{-(7/5)} \epsilon^{-(3/5)} p^{3/20} + 3 \cdot 2^{-(1/5)} \epsilon^{1/5} p^{-(1/20)})^4.$$

This can be shown as follows. The first two terms of both sides are equal. Furthermore

$$3 \cdot 2^{4/5} \epsilon^{-(4/5)} p^{1/5} < 3 \cdot 2^{4/5} \epsilon^{-(4/5)} p^{1/5} \cdot \frac{9}{8} = 27 \cdot 2^{-(11/5)} \epsilon^{-(4/5)} p^{1/5}$$

and

$$16 < 27 + 81 \cdot 2^{-(4/5)} \epsilon^{4/5} p^{-(1/5)}.$$

Consequently (74) holds. From (73) and (74), we obtain

$$u < \text{Max} \{ 2^{3/5} \epsilon^{2/5} p^{2/5} + 2^{4/5} \epsilon^{1/5} p^{1/5} + 1, \\ 4\epsilon p^{1/4} (2^{-(7/5)} \epsilon^{-(3/5)} p^{3/20} + 3 \cdot 2^{-(1/5)} \epsilon^{1/5} p^{-(1/20)}) + 2^{-(6/5)} \epsilon^{1/5} p^{1/5} + 4\epsilon - 1 \}, \\ u < \text{Max} \{ 2^{3/5} \epsilon^{2/5} p^{2/5} + 2^{4/5} \epsilon^{1/5} p^{1/5} + 1, \\ 2^{3/5} \epsilon^{2/5} p^{2/5} + 2^{9/5} \cdot 3\epsilon^{8/5} p^{1/5} + 2^{-(6/5)} \epsilon^{1/5} p^{1/5} + 4\epsilon - 1 \}.$$

Hence

$$u < 2^{3/5} \epsilon^{2/5} p^{2/5} + 2^{9/5} \epsilon^{8/5} p^{1/5} (3 + 2^{-3} \epsilon^{-1}) + 4\epsilon - 1.$$

This gives a contradiction to (71), showing that (71) can not hold. Hence, according to (66)

$$u < 2^{3/5} p^{2/5} + 2^{-(6/5)} \cdot 25 \cdot p^{1/5} + 3 \text{ for } p = 8n + 5 > 2048, \\ u < 2p^{2/5} + 4\frac{9}{2} p^{1/5} + 7 \text{ for } p = 8n + 3.$$

This proves the statement (42) for all primes of the form $8n + 3$ and for primes of the form $8n + 5$ which exceed 2048.

It remains to prove (42) for primes of the form $8n + 5$ which are less than 2048. For these primes, we have, according to (15)

$$u < \sqrt{p+4} + 2.$$

It is therefore sufficient to show that for $p < 2048$ we have

$$(75) \quad \sqrt{p+4} + 2 < 2^{3/5} p^{2/5} + 2^{-(6/5)} \cdot 25 p^{1/5} + 3.$$

But this follows from

$$p + 4 < (2^{3/5} p^{2/5} + 2^{-(6/5)} \cdot 25 p^{1/5})^2 = 2^{6/5} p^{4/5} + 2^{2/5} \cdot 25 p^{3/5} + 2^{-(12/5)} \cdot 625 p^{2/5}$$

which is true, since $p < 2048$ and therefore $p < 25p^{3/5}$. Hence (75) holds and Theorem 8 is proved.

It will be necessary to improve Theorem 8 still further. We have

THEOREM 9. *Let p be a prime of the form $8n + 5$ which satisfies*

$$(76) \quad p > 6\frac{4}{81} \cdot 10^5.$$

Assume that the two least quadratic non-residues u, v which are odd primes satisfy the condition

$$(77) \quad u < v < 6u.$$

If we set

$$(78) \quad \rho = v/u + 1,$$

we have

$$(79) \quad u < 2(p/\rho)^{2/5} + (24/\rho + \frac{1}{2})(p/\rho)^{1/5} + 9/\rho.$$

Proof. From (77) and (78) we obtain

$$(80) \quad 2 < \rho < 7.$$

For $u < 16$, the statement (79) is certainly true because of (76). For $u > 16$, we have

$$u - (2\rho - 2) > 4$$

according to (80). Let z be an arbitrary odd integer which satisfies

$$(81) \quad 2\rho - 2 \leq 2z < u.$$

We consider the numbers of the form $p \pm h_v u$ with integral $h_v \geq 0$ which lie in the interval $\{p - v + 1, \dots, p + v - 1\}$. We have here $h_v u < v$, and according to (78)

$$(82) \quad h_v < \rho - 1.$$

At most one of the numbers $p \pm h_v u$ is divisible by $2z$. Indeed, if two of these numbers were divisible by $2z$, so would their difference, which is of the form $h^* u$ with $|h^*| < 2\rho - 2$, because of (82). But u is a prime and therefore prime to $2z$, according to (81). Hence h^* would be divisible by $2z$. This, however, is impossible because (81) implies

$$|h^*| < 2\rho - 2 \leq 2z.$$

Consequently, either all the numbers $p \pm h_v u$ in the interval $\{p - v + 1 \dots p\}$, or all those numbers in the interval $\{p \dots p + v - 1\}$ are not divisible by $2z$. Let us assume that $\{p \dots p + v - 1\}$ does not contain a number $p \pm h_v u$ which is divisible by $2z$. In the other case we may argue in exactly the same manner.

Let V be the interval $\{p - u + 1 \dots p + v - 1\}$, and suppose that

$$(83) \quad k \cdot 2z, (k + 1)2z, \dots, (k + l - 1)2z$$

are the multiples of $2z$ in V . Since it has been assumed that the interval $\{p \dots p + v - 1\}$ does not contain a multiple of $2z$ of the form $p \pm h_v u$, and since the interval $\{p - u + 1 \dots p - 1\}$ does not contain a number of the form $p \pm h_v u$, none of the numbers (83) is of the form $p \pm h_v u$ with integral h_v . However, all the even numbers of V , except the numbers $p \pm h_v u$, are quadratic residues (mod p). Since 2 is a non-residue and z is a residue, according to (81), the numbers

$$(84) \quad k, k + 1, \dots, k + l - 1$$

form a sequence of l quadratic non-residues, analogously to (46). From (83) it follows that (45) again holds.

Let a , A , and s have the same significance as in the proof of Theorem 8. We see in the same manner as above, that (48) and (64) hold. Then

$$(85) \quad a < \left(\frac{p}{2z}\right)^{\frac{1}{2}},$$

$$(86) \quad s \leq 4[\sqrt{2a}] + 2.$$

If now

$$(87) \quad u \geq \text{Max} \left\{ a + 2 + [\sqrt{2a}], \frac{2zs + 1}{\rho} \right\},$$

then we could show as above that A cannot contain a sequence of s non-residues.

Because of (83), V contains exactly l integral multiples of $2z$. On the other hand, V contains exactly $u + v - 1$ consecutive integers, and therefore at least $\left[\frac{u + v - 1}{2z} \right]$ multiples of $2z$. Consequently, we have

$$l \geq \left[\frac{u + v - 1}{2z} \right] = \left[\frac{u\rho - 1}{2z} \right] \geq \left[\frac{2zs + 1 - 1}{2z} \right] = s.$$

because of (78) and (87), and there lies a sequence $k, k + 1, \dots, k + l - 1$ of at least s non-residues in A , according to (84). This gives a contradiction to the result above, and hence (87) is not true. We have, therefore

$$u < \text{Max} \left\{ a + 2 + [\sqrt{2a}], \frac{2zs + 1}{\rho} \right\}.$$

Since the first expression on the right-hand side is an integer, we see, because of (86) and (85), that

$$\begin{aligned} u &\leq \text{Max} \left\{ a + 1 + [\sqrt{2a}], \frac{2z(4[\sqrt{2a}] + 2) + 1}{\rho} \right\}, \\ u &< \text{Max} \left\{ \sqrt{\frac{p}{2z}} + \sqrt[4]{\frac{2p}{z}} + 1, \frac{2z}{\rho} \left(4\sqrt[4]{\frac{2p}{z}} + 2 \right) + \frac{1}{\rho} \right\}, \\ (88) \quad u &< \text{Max} \left\{ \sqrt{\frac{p}{2z}} + \sqrt[4]{\frac{2p}{z}} + 1, \frac{8}{\rho} \sqrt[4]{2pz^3} + \frac{4z + 1}{\rho} \right\}. \end{aligned}$$

Let us assume that

$$(89) \quad u \geq 2 \left(\frac{p}{\rho} \right)^{2/5} + \left(\frac{24}{\rho} + \frac{1}{2} \right) \left(\frac{p}{\rho} \right)^{1/5} + \frac{9}{\rho}.$$

We then determine the odd integer z_0 such that

$$(90) \quad \frac{1}{8} \rho^{4/5} p^{1/5} + 2 > z_0 > \frac{1}{8} \rho^{4/5} p^{1/5}.$$

Because of (80), we have $\rho < 7$. From (76) it follows that

$$(91) \quad p > \frac{2^{11} \cdot 5^5}{3^4} = \frac{2^{15} \cdot 5^5}{6^4} \geq \frac{2^{15}[\rho - 1]^5}{[\rho]^4} \geq \frac{2^{15}[\rho - 1]^5}{\rho^4}$$

since $\frac{(x-1)^5}{x^4}$ is increasing with x for $x \geq 1$. From (90) and (91) we obtain

$$\begin{aligned} z_0 &> \frac{1}{8} \rho^{4/5} p^{1/5} > [\rho - 1], \\ (92) \quad z_0 &\geq [\rho] > \rho - 1, \end{aligned}$$

since z_0 was an integer

Furthermore, we have to show that

$$(93) \quad 2z_0 < u.$$

But this is true, since

$$2z_0 < \frac{1}{4} \rho \left(\frac{p}{\rho} \right)^{1/5} + 4 < 2 \left(\frac{p}{\rho} \right)^{1/5} + 4 < 2 \left(\frac{p}{\rho} \right)^{1/5} + \frac{24}{\rho} + \frac{9}{\rho} < 2 \left(\frac{p}{\rho} \right)^{2/5} + \left(\frac{24}{\rho} + \frac{1}{2} \right) \left(\frac{p}{\rho} \right)^{1/5} + \frac{9}{\rho} \leq u$$

because of (90), (80), (76), and (89).

According to (92) and (93) the conditions (81) are satisfied for $z = z_0$. Hence it follows from (88) and (90) that

$$(94) \quad u < \text{Max} \left\{ \left(\frac{4p}{\rho^{4/5} p^{1/5}} \right)^{1/2} + \left(\frac{16p}{\rho^{4/5} p^{1/5}} \right)^{1/4} + 1, \right. \\ \left. \frac{8}{\rho} p^{1/4} \left(\frac{1}{256} \rho^{12/5} p^{3/5} + \frac{3}{16} \rho^{8/5} p^{2/5} + 3\rho^{4/5} p^{1/5} + 16 \right)^{1/4} + \frac{1}{2} \left(\frac{p}{\rho} \right)^{1/5} + \frac{9}{\rho} \right\}.$$

We now state that

$$(95) \quad \frac{1}{256} \rho^{12/5} p^{3/5} + \frac{3}{16} \rho^{8/5} p^{2/5} + 3\rho^{4/5} p^{1/5} + 16 < \frac{1}{256} \rho^{12/5} p^{3/5} + \frac{3}{16} \rho^{8/5} p^{2/5} \\ + \frac{27}{8} \rho^{4/5} p^{1/5} + 27 + 81\rho^{-(4/5)} p^{-(1/5)} = \left(\frac{1}{4} \rho^{3/5} p^{3/20} + 3\rho^{-(1/5)} p^{-(1/20)} \right)^4.$$

This follows immediately from $3 < 27/8$ and $16 < 27$, since the first two terms on either side of the inequality are equal. We obtain from (94) and (95)

$$u < \text{Max} \left\{ 2 \left(\frac{p}{\rho} \right)^{2/5} + 2 \left(\frac{p}{\rho} \right)^{1/5} + 1, \right. \\ \left. \frac{8}{\rho} p^{1/4} \left(\frac{1}{4} \rho^{3/5} p^{3/20} + 3\rho^{-(1/5)} p^{-(1/20)} \right) + \frac{1}{2} \left(\frac{p}{\rho} \right)^{1/5} + \frac{9}{\rho} \right\}, \\ u < \text{Max} \left\{ 2 \left(\frac{p}{\rho} \right)^{2/5} + 2 \left(\frac{p}{\rho} \right)^{1/5} + 1, 2 \left(\frac{p}{\rho} \right)^{2/5} + \left(\frac{24}{\rho} + \frac{1}{2} \right) \left(\frac{p}{\rho} \right)^{1/5} + \frac{9}{\rho} \right\}.$$

Since $\rho < 7$, each term of the second expression on the right side is not smaller than the corresponding term of the first expression. Hence

$$u < 2 \left(\frac{p}{\rho} \right)^{2/5} + \left(\frac{24}{\rho} + \frac{1}{2} \right) \left(\frac{p}{\rho} \right)^{1/5} + \frac{9}{\rho}.$$

This gives a contradiction to (89) which cannot then be true, proving (79).

5. Proof of the non-existence of the algorithm for $p > 3\,300\,000$.

Based on the results of the preceding section we shall now show that there does not exist an Euclidean algorithm in $P(p^{1/2})$, if p is a prime of the form $24n + 13$, and $p > 3\,300\,000$.

According to Theorem 6, it is sufficient to assume $v < 6u$, and hence $2 < \rho < 7$ because of (80). Theorem 9 gives then the inequality (79) for u .

According to Theorem 2, the algorithm certainly does not exist, if

$$(96) \quad 3uv = 3(\rho - 1)u^2 < p$$

or because of (79)

$$(97) \quad 3(\rho-1)u^2 < 3(\rho-1) \left\{ 2\left(\frac{p}{\rho}\right)^{2/5} + \left(\frac{24}{\rho} + \frac{1}{2}\right)\left(\frac{p}{\rho}\right)^{1/5} + \frac{9}{\rho} \right\}^2 < p.$$

We now state that the function of ρ

$$(98) \quad \phi(\rho) = 3 \left\{ 2\left(\frac{p}{\rho}\right)^{2/5} (\rho-1)^{1/2} + \frac{24}{\rho} \left(\frac{p}{\rho}\right)^{1/5} (\rho-1)^{1/2} + \frac{1}{2} \left(\frac{p}{\rho}\right)^{1/5} (\rho-1)^{1/2} + \frac{9}{\rho} (\rho-1)^{1/2} \right\}^2$$

is increasing in the interval $2 < \rho \leq 7$, if $p > 3\,300\,000$ is fixed. Since $\frac{1}{2}(p/\rho)^{1/5}(\rho-1)^{1/2}$ is increasing for $\rho > 2$, it is sufficient to show that

$$\psi(\rho) = (\rho-1)^{1/2} \left\{ 2\left(\frac{p}{\rho}\right)^{2/5} + \frac{24}{\rho} \left(\frac{p}{\rho}\right)^{1/5} + \frac{9}{\rho} \right\}$$

is increasing for $2 < \rho \leq 7$, if $p > 3\,300\,000$ is fixed. Differentiating with regard to ρ for a fixed p and setting $p^{1/5} = y$, we obtain

$$\begin{aligned} \psi'(\rho) &= \frac{1}{2} (\rho-1)^{-(1/2)} \{ 2y^2 \rho^{-(2/5)} + 24y \rho^{-(6/5)} + 9\rho^{-1} \} \\ &\quad - (\rho-1)^{1/2} \left\{ \frac{4}{5} y^2 \rho^{-(7/5)} + \frac{144}{5} y \rho^{-(11/5)} + 9\rho^{-2} \right\}, \\ (99) \quad 10(\rho-1)^{1/2} \rho^{11/5} \psi'(\rho) &= y^2 \rho^{4/5} (2\rho+8) + y(288-168\rho) + 45\rho^{1/5} (2-\rho). \end{aligned}$$

We have to distinguish between two cases

$$1) \quad 2 < \rho \leq 3.$$

Because of $p > 3\,300\,000$ we have $y > 20$ and therefore from (99)

$$(100) \quad 10(\rho-1)^{1/2} \rho^{11/5} \psi'(\rho) > y(40\rho+160+288-168\rho) + 45 \cdot 3^{1/5} (2-\rho) > y(448-128\rho) + 90(2-\rho) > 64y-90 > 0.$$

$$2) \quad 3 < \rho \leq 7.$$

From (99) we obtain

$$\begin{aligned} (101) \quad 10(\rho-1)^{1/2} \rho^{11/5} \psi'(\rho) &> 3^{4/5} y^2 (2\rho+8) + y(288-168\rho) \\ &\quad + 45 \cdot 7^{1/5} (2-\rho) > 2.2y^2 (2\rho+8) + y(288-168\rho) - 450 \\ &= y(88\rho+352+288-168\rho) - 450 > 80y-450 > 0. \end{aligned}$$

Because of (100) and (101) we have $\psi'(\rho) > 0$ for $2 < \rho \leq 7$, hence $\psi(\rho)$ and $\phi(\rho)$ are increasing in this interval, if $p > 3\,300\,000$ is fixed. It then follows from (98) for $2 < \rho \leq 7$ that

$$\begin{aligned} (102) \quad \phi(\rho) &< 72 \left(\frac{p}{7}\right)^{4/5} + \left(\frac{1728}{7} + 36\right) \left(\frac{p}{7}\right)^{3/5} \\ &\quad + \left(\frac{10368}{49} + \frac{1080}{7} + \frac{9}{2}\right) \left(\frac{p}{7}\right)^{2/5} + \left(\frac{7776}{49} + \frac{162}{7}\right) \left(\frac{p}{7}\right)^{1/5} + \frac{1458}{49}. \end{aligned}$$

According to (96), (97), (98), and (102), the Euclidean algorithm does not exist in the field $P(p^{1/2})$ if

$$3uv = 3(p-1)u^2 < \phi(p) < 72\left(\frac{p}{7}\right)^{4/5} + \frac{1980}{7}\left(\frac{p}{7}\right)^{3/5} \\ + \frac{36297}{98}\left(\frac{p}{7}\right)^{2/5} + \frac{8910}{49}\left(\frac{p}{7}\right)^{1/5} + \frac{1458}{49} < p.$$

This is true for $p > 3\,300\,000$, since the polynomial

$$7x^5 - 72x^4 - \frac{1980}{7}x^3 - \frac{36297}{98}x^2 - \frac{8910}{49}x - \frac{1458}{49}$$

has only one positive root, and assumes a positive value for $x = \left(\frac{3\,300\,000}{7}\right)^{1/5}$, as can be seen by a simple computation.

6. The case $p < 3\,300\,000$. The primes $p < 3\,300\,000$ must be investigated directly using the Theorems 5, 7, 1, and 2. This investigation is tedious but the methods are straightforward.

If the integer 5 is a non-residue for such a p , then we cannot have an algorithm, according to Theorem 5, if $p > 24 \cdot 5^2 = 600$. Analogously, we can argue for $p > 24 \cdot 7^2$, if $(7/p) = -1$, for $p > 24 \cdot 11^2$, if $(11/p) = -1$, and for $p > 24 \cdot 13^2$, if $(13/p) = -1$.

We can then form the 180 arithmetical progressions which contain those primes $p = 24m + 13$, for which 5, 7, 11, and 13 are residues. We consider those primes $p < 3\,300\,000$ using the modules 17, 19, 23, . . . as far as it is necessary for the construction of the least quadratic non-residue. For most of these p , it follows from Theorems 5, 7 or 2 that we have no Euclidean algorithm in $P(p^{1/2})$. The only exceptional cases are for $p = 13, 37, 61, 109, 181, 229$, and 421. For $p = 13$ and $p = 37$, the algorithm exists, as has been mentioned in the introduction. For $p = 181, 229$, and 421, there is no algorithm, as follows from Theorem 1, because of

$$181 = 7 \cdot 17 + 2 \cdot 31, \quad 229 = 7 \cdot 13 + 6 \cdot 23, \quad 421 = 13 \cdot 19 + 6 \cdot 29.$$

Whether or not the algorithm exists for $p = 61$ and $p = 109$, I cannot decide.

We have thus the result

THEOREM 10. *There is no Euclidean algorithm in the field $P(p^{1/2})$, if $p > 109$ is a prime of the form $24n + 13$.*

THEOREM 11. *Let $p > 421$ be a prime of the form $24n + 13$, and u, v the two least quadratic non-residues (mod p), which are odd primes. Then we have*

$$3uv < p.$$

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

POSTULATIONAL BASES FOR THE UMBRAL CALCULUS.*

By E. T. BELL.

As the somewhat condensed treatment of the umbral calculus which I gave elsewhere¹ has been misunderstood² a fuller treatment than was given before is desirable. Incidentally, what follows validates the purely formal uses of this calculus, or of its special cases, which have appeared in the literature, when such uses give correct results. There are immediate generalizations to abstract commutative rings, obtainable by obvious modifications of the following; but as such generalizations seem to be of no use at present, it seems hardly worth while to develop them.

1. Rational operations on umbrae.

(1.1) Real, or complex, numbers are called *scalars*. The sign \equiv denotes either definitional identity or identity as in algebra; which, will be clear from the context.

(1.2) Scalars are denoted by small Latin letters *with non-negative integer suffixes*, thus x_N ($N = 0, 1, \dots$), or by small Greek letters, α, β, \dots . As usual, the sum, product of any scalars α, β are $\alpha + \beta, \alpha\beta$, and 0, 1 have their usual meanings.

(1.3) Latin capitals, A, \dots, N, \dots denote non-negative integers.

(1.4) If x_N ($N = 0, 1, \dots$) are any scalars, the one-rowed matrix $(x_0, x_1, \dots, x_N, \dots)$ is denoted by x : $x \equiv (x_0, x_1, \dots, x_N, \dots)$.

(1.5) The $(N + 1)$ -th element, $N = 0, 1, \dots$, of x in (1.4) is denoted by x^N :

$$x^N \equiv x_N \qquad (N = 0, 1, \dots).$$

(1.6) The x in (1.4) is called an *umbra*; x is the umbra of $(x_0, x_1, \dots, x_N, \dots)$, or of the sequence x_N ($N = 0, 1, \dots$). Note that an umbra has neither exponent nor suffix.

(1.7) *Equality* of umbrae is matrix equality: if x is as in (1.4), and

* Received April 8, 1940.

¹ "Algebraic arithmetic," *American Mathematical Society Publications*, vol. 7 (1927), pp. 146-159.

² G. Temple, *Journal of the London Mathematical Society*, vol. 12 (1937), p. 114. Professor Temple has seen the present note, and writes (Feb. 21, 1938) that it clears up the obscurity.

$y \equiv (y_0, y_1, \dots, y_N, \dots)$, 'x is equal to y,' written $x \doteq y$, if, and only if, $x_N = y_N$ ($N = 0, 1, \dots$). Hence

$$(1.71) \quad x \doteq x.$$

$$(1.72) \quad \text{If } x \doteq y, \text{ then } y \doteq x.$$

$$(1.73) \quad \text{If } x \doteq y, \text{ and } y \doteq z, \text{ then } x \doteq z.$$

(1.8) The coefficient of $x_1^{s_1} \dots x_T^{s_T}$ in the expansion of $(x_1 + \dots + x_T)^N$ by the multinomial theorem, is denoted by M_{s_1, \dots, s_T} . Note that exponents and suffixes 0, 1 are to be indicated precisely in the same way as exponents and suffixes > 1 .

The next refer to rational functions of umbrae, and define 'umbral scalar multiplication,' 'umbral addition,' etc. The qualification 'umbral' will be dropped, as it is taken care of in the notation.

(1.9) The scalar product, αx , of α and $x \equiv (x_0, \dots, x_N, \dots)$ is

$$\alpha x \equiv \alpha(x_0, \dots, x_N, \dots) \equiv (\alpha x_0, \dots, \alpha x_N, \dots).$$

By definition, $\alpha x = \alpha x$.

Now αx is an umbra, by (1.6), and it is a compound symbol. To denote the $(N+1)$ -th element of αx in accordance with (1.5), we write $\{\alpha x\}^N$; thus

$$(1.91) \quad \{\alpha x\}^N \equiv \alpha x^N \equiv \alpha x_N.$$

Similarly, if $*$ is any compound symbol of scalars and umbrae, and if $*$ is an umbra, the $(N+1)$ -th element of $*$ is denoted by $\{*\}^N$.

(1.10) The sum, s , $s \equiv \alpha a \dot{+} \dots \dot{+} \xi x$, of $\alpha a, \dots, \xi x$, where

$$a \equiv (a_0, \dots, a_N, \dots), \dots, x \equiv (x_0, \dots, x_N, \dots),$$

is

$$s \equiv (\alpha a_0 + \dots + \xi x_0, \dots, \alpha a_N + \dots + \xi x_N, \dots).$$

Hence

$$(1.101) \quad \{\alpha a \dot{+} \dots \dot{+} \xi x\}^N = \alpha a_N + \dots + \xi x_N;$$

(1.102) Addition, $\dot{+}$, of umbrae is commutative and associative;

(1.103) There is a unique z , the zero umbra, such that $x \dot{+} z \doteq x$ for every x :

$$z \equiv (0, \dots, 0, \dots);$$

(1.104) For every x there is a unique y such that $x \dot{+} y \doteq z$; y is called the negative of x ; $y \equiv (-1)x$, and is denoted by $\dot{-}x$;

(1.105) With respect to $\dot{+}$ the set of all umbrae is an abelian group; the inverse of x in the group is $\dot{-}x$, and the identity of the group is z .

(1.11) If no two of a, \dots, x are equal as defined in (1.7), a, \dots, x are said to be *distinct*. In (1.12)–(1.125), a, \dots, x are distinct.

(1.12) If a, \dots, x are T distinct umbrae, $(\alpha a \dot{+} \dots \dot{+} \xi x)^N$ denotes the scalar p_N ,

$$(1.120) \quad p_N \equiv (\alpha a \dot{+} \dots \dot{+} \xi x)^N \equiv \Sigma M_{s_1, \dots, s_T} \alpha^{s_1} \dots \xi^{s_T} a^{s_1} \dots x^{s_T},$$

(see (1.8)). In particular,

$$(1.121) \quad p_0 = a_0 \dots x_0.$$

Hence, by (1.5),

$$(1.122) \quad (\alpha a \dot{+} \dots \dot{+} \xi x)^N = \Sigma M_{s_1, \dots, s_T} \alpha^{s_1} \dots \xi^{s_T} a^{s_1} \dots x^{s_T}$$

the left of which is called the N -th power of the sum $\alpha a \dot{+} \dots \dot{+} \xi x$. Hence such powers are expanded by the multinomial theorem, and $\dot{+}$ is replaced by $+$ in the result.

If p_N is as above defined, and $p \equiv (p_0, \dots, p_N, \dots)$, then $p^N = p_N$, by (1.5). Note the distinction, as shown in (1.101), (1.122), between

$$\{\alpha a \dot{+} \dots \dot{+} \xi x\}^N, \quad (\alpha a \dot{+} \dots \dot{+} \xi x)^N,$$

only the second of which is a power; both are scalars.

By (1.121),

$$(1.123) \quad (\alpha a \dot{+} \dots \dot{+} \xi x)^0 = a_0 \dots x_0.$$

In (1.122) replace N by $N + R$. The resulting scalar,

$$(\alpha a \dot{+} \dots \dot{+} \xi x)^{N+R},$$

is called the *product*,

$$(\alpha a \dot{+} \dots \dot{+} \xi x)^N \cdot (\alpha a \dot{+} \dots \dot{+} \xi x)^R,$$

of

$$(\alpha a \dot{+} \dots \dot{+} \xi x)^N, \quad (\alpha a \dot{+} \dots \dot{+} \xi x)^R;$$

$$(1.124) \quad (\alpha a \dot{+} \dots \dot{+} \xi x)^N \cdot (\alpha a \dot{+} \dots \dot{+} \xi x)^R \equiv (\alpha a \dot{+} \dots \dot{+} \xi x)^{N+R}.$$

It follows that this multiplication, \cdot , is commutative and associative, and that it has the 'identity' $(\alpha a \dot{+} \dots \dot{+} \xi x)^0$. The right of (1.124) may be (and is) calculated from the left by expanding each of the factors $(\)^N, (\)^R$ by the multinomial theorem, multiplying the resulting (scalar) polynomials together as in common algebra and finally degrading all exponents of small Latin letters to suffixes. For example, noting that $\alpha^0 = \beta^0 = 1$, and $\alpha^1 = \alpha$, $\beta^1 = \beta$, since α, β are scalars, we have

$$\begin{aligned} & (\alpha a \dot{+} \beta b)^1 \cdot (\alpha a \dot{+} \beta b)^2 \\ &= (\alpha a^1 b^0 + \beta a^0 b^1) \cdot (\alpha^2 a^2 b^0 + 2\alpha\beta a^1 b^1 + \beta^2 a^0 b^2) \end{aligned}$$

$$\begin{aligned}
&= \alpha^3 a^3 b^0 + 3\alpha^2 \beta a^2 b^1 + 3\alpha \beta^2 a^1 b^2 + \beta^3 a^0 b^3, \\
&= \alpha^3 a_3 b_0 + 3\alpha^2 \beta a_2 b_1 + 3\alpha \beta^2 a_1 b_2 + \beta^3 a_0 b_3, \\
&= (\alpha a + \beta b)^3.
\end{aligned}$$

As a mere convenience of notation we write

$$\begin{aligned}
(1.125) \quad &(\xi x)^N \cdot [(\alpha a)^M + (\beta b)^R + \cdots + (\gamma c)^S] \\
&\equiv (\xi x)^N \cdot (\alpha a)^M + (\xi x)^N \cdot (\beta b)^R + \cdots + (\xi x)^N \cdot (\gamma c)^S,
\end{aligned}$$

the (scalar) sum of scalars on the right defining the expression on the left. Similarly for an infinity of scalar summands.

All in this section (1.12) refers only to the case in which the T umbrae a, \cdots, x are distinct. The contrary case is equally important in applications of the calculus, and requires special consideration.

(1.13) If in $\alpha x + \cdots + \xi x$ there are precisely T summands $\alpha x, \cdots, \xi x$, each of which is a scalar product of a scalar and x , we replace (\rightarrow) the T x 's by T distinct umbrae, say a, \cdots, x , in any order, and indicate this replacement by writing

$$(1.131) \quad \alpha x + \cdots + \xi x \rightarrow \alpha a + \cdots + \xi x.$$

Then $(\alpha a + \cdots + \xi x)^N$ is to be calculated by (1.122), and the exponents are degraded, as in (1.120). In the result, each of a, \cdots, x is replaced (\leftarrow) by x ; the resulting polynomial is defined to be N -th power $(\alpha x + \cdots + \xi x)^N$ of the sum $\alpha x + \cdots + \xi x$.

For example,

$$\begin{aligned}
(\alpha x + \beta x)^3 &\rightarrow (\alpha a + \beta x)^3; \\
(\alpha a + \beta x)^3 &= \alpha^3 a_3 x_0 + 3\alpha^2 \beta a_2 x_1 + 3\alpha \beta^2 a_1 x_2 + \beta^3 a_0 x_3, \\
&\leftarrow \alpha^3 x_3 x_0 + 3\alpha^2 \beta x_2 x_1 + 3\alpha \beta^2 x_1 x_2 + \beta^3 x_0 x_3; \\
(\alpha x + \beta x)^3 &= (\alpha^3 + \beta^3) x_0 x_3 + 3\alpha \beta (\alpha + \beta) x_1 x_2.
\end{aligned}$$

The relation (1.124) holds also for powers $(\alpha x + \cdots + \xi x)^N$ when therein the replacements \rightleftharpoons are made.

Similarly, if in a $(+)$ sum s there are precisely A summands each of which is a scalar product of a scalar and x, \cdots , precisely C summands each of which is a scalar product of a scalar and w , and if these summands exhaust s , the $S \equiv A + \cdots + C$ x 's, w 's, are replaced (\rightarrow) by S distinct umbrae, say $s \rightarrow t$. Then $(t)^N$ is calculated by (1.122), (1.120), and the final replacement (\leftarrow) of the S distinct umbrae by those introduced by (\rightarrow). These powers $(s)^N$ also satisfy (1.124).

(1.132) Hence (1.124) holds for any umbrae a, \cdots, x , distinct or not.

(1.14) x^N was defined in (1.5); it denotes the scalar x_N . Hence, since

multiplication of scalars is indicated (as always) by mere juxtaposition, without any symbol denoting the operation of multiplication,

$$(1.141) \quad x^N x^R = x_N x_R.$$

Since this multiplication is multiplication of scalars, it is commutative and associative.

In $(\alpha a + \cdots + \xi x)^N$, defined in (1.120), take $\xi = 1$ and each of the other scalars $= 0$. Then by (1.9), (1.103), $(0a + \cdots + 1x)^N = (x)^N$. Note that $()$ is not omitted on the right. By (1.120), $(x)^N = x^N$. Hence, by (1.5), $(x)^N = x_N$. By (1.124), $(x)^N \cdot (x)^R = (x)^{N+R}$, and hence, by what has just been shown, $x^N \cdot x^R = x^{N+R} = x_{N+R}$,

$$(1.142) \quad x_N \cdot x_R = x_{N+R}.$$

Thus, unless $x_N x_R = x_{N+R}$, $x_N x_R \neq x_N \cdot x_R$. The 'dot multiplication,' \cdot , is an operation peculiar to the calculus, and will be explicitly indicated where there is any possibility of confusion.

Similarly, $(\alpha a + \cdots + \xi x)^N (\alpha x + \cdots + \xi x)^R$, without the dot, is the (scalar) product of the scalars $(\alpha a + \cdots + \xi x)^N$, $(\alpha x + \cdots + \xi x)^R$, which are defined in (1.222); and this scalar product is different from the dot product in (1.124). To see the difference in an example, we compare the example illustrating (1.124) with the following:

$$\begin{aligned} & (\alpha a + \beta b)^1 (\alpha a + \beta b)^2, \\ &= (\alpha a_1 b_0 + \beta a_0 b_1) (\alpha^2 a_2 b_0 + 2\alpha\beta a_1 b_1 + \beta^2 a_0 b_2), \\ &= \alpha^3 a_1 a_2 b_0^2 + \alpha^2 \beta b_0 b_1 (2a_1^2 + a_0 a_2) + \alpha\beta^2 a_0 a_1 (2b_1^2 + b_0 b_2) + \beta^3 a_0^2 b_1 b_2, \\ &\neq (\alpha a + \beta b)^1 \cdot (\alpha a + \beta b)^2. \end{aligned}$$

(1.15) A particular case of (1.120) occurs so frequently that a special notation is convenient. If $s \equiv \alpha x + \cdots + \alpha x$ is a sum of precisely A scalar products αx , we write

$$(1.151) \quad A \cdot \alpha x \equiv s \equiv \alpha x + \cdots + \alpha x.$$

There can be no confusion between the dot in $A \cdot \alpha x$ and that in (1.124), since here the dot is between a scalar and an umbra, while in (1.124) it is between two scalars. If desired, the dot in (1.151) may be circled, thus \odot . It would be incorrect to write $A\alpha x$ instead of $A \cdot \alpha x$, since $A\alpha$ is a scalar, and hence, by (1.9), $A\alpha x$ is a scalar product.

(1.16) *Umbra* multiplication can be defined in many (actually, an infinity of) ways to yield algebras simply isomorphic with parts of the common algebra of scalars, for example rings. Here we need mention only that species of umbral multiplication which is directly applicable to the power series in § 2. It will not be used in the sequel.

$$(1.161) \quad x \equiv (x_0/0!, x_1/1!, \dots, x_N/N!, \dots)$$

is said to be of *e-type* ($e \equiv$ 'exponential'). Hence, if y is of *e-type*, $y^N = y_N/N!$. If w is not of *e-type*, it is replaced by \bar{w} , in which $w^N \equiv \bar{w}^N/N!$, until after all calculations involving \bar{w} have been completed, when \bar{w}_N is replaced by $N!w_N$.

Let $x \equiv (x_0/0!, \dots, x_N/N!, \dots)$, $y \equiv (y_0/0!, \dots, y_N/N!, \dots)$ be of *e-type*. The product, xy , of x, y (in this order) is the matrix p which is such that

$$(1.162) \quad p^N \equiv \frac{(x + y)^N}{N!};$$

$$(1.163) \quad xy \equiv \left(\frac{(x + y)^0}{0!}, \frac{(x + y)^1}{1!}, \dots, \frac{(x + y)^N}{N!}, \dots \right).$$

Hence umbral multiplication is commutative and associative. Thus powers may be defined as usual; the A -th power of x is denoted by $x^{(A)}$, to distinguish it from x^A .

2. Power series. The set of all (formal) power series in the variable θ is closed under the four rational operations. Division is immediately referred to multiplication, and need not be separately discussed. Irrational functions of these power series also occur, but as they are of less interest than the rational functions, and are readily investigated if desired, they will not be considered here. The use of formal (disregard of convergence) power series can be justified in detail, if not obviously legitimate in the present connection (for example, as in my paper, *Transactions of the American Mathematical Society*, vol. 25, 1923, 135-54); however, there is sufficient generality in the set of all power series in θ convergent in the same domain $|\theta| > 0$ to show here how the definitions, etc., in § 1 give immediately the algorithms of Blissard's umbral calculus.

If $x \equiv (x_0, \dots, x_N, \dots)$ we write

$$(2.1) \quad e^{x\theta} \equiv \sum_{N=0}^{\infty} x_N (\theta^N/N!),$$

where e has its usual meaning (2.7). Thus, by (1.5),

$$(2.11) \quad \xi e^{x\theta} = \xi \sum_{N=0}^{\infty} x_N (\theta^N/N!).$$

By either of these, $\xi e^{x\theta}$ is a scalar. Hence if $\Lambda(\xi, \dots, \eta)$ is a polynomial in ξ, \dots, η with scalar coefficients, $\Lambda \equiv \Lambda(\xi e^{x\theta}, \dots, \eta e^{y\theta})$ is a scalar, as is also the N -th derivative $\partial_\theta^N \Lambda$ of Λ with respect to θ . By writing Λ as a MacLaurin

series in θ , we express it in the form $\tau e^{w\theta}$, and similarly for the derivative. For any Λ (or its derivative) the appropriate $\tau e^{w\theta}$ is built up by repeated applications of the elementary identities (2.2)–(2.4) in θ .

$$(2.2) \quad e^{x\theta} e^{y\theta} = e^{(x+y)\theta} \equiv \sum_0^\infty (x + y)^N (\theta^N / N!),$$

which, by (1.120), is merely the formal multiplication of two MacLaurin series to produce a third. Generally, for any number of factors on the left,

$$(2.21) \quad e^{\xi x\theta} \cdots e^{\eta y\theta} \equiv e^{(\xi x + \cdots + \eta y)\theta} \equiv \sum_0^\infty (\xi x + \cdots + \eta y)^N (\theta^N / N!).$$

For addition, (1.101) gives

$$(2.2) \quad e^{x\theta} + e^{y\theta} \equiv e^{\{x+y\}\theta} \equiv \sum_0^\infty \{x + y\}^N (\theta^N / N!),$$

with the obvious extension to any number of summands.

Powers are obtained directly from (2.21), or more conveniently thence by (1.151):

$$(2.3) \quad [e^{\xi x\theta}]^A \equiv e^{(A \cdot \xi x)\theta} \equiv \sum_0^\infty (A \cdot \xi x)^N (\theta^N / N!).$$

For derivation, we have

$$\begin{aligned} \partial_\theta^N e^{\xi x\theta} &\equiv \partial_\theta^N \sum_{M=0}^\infty \xi^M x_M (\theta^M / M!), \\ &= \sum_{M=0}^\infty \xi^{N+M} x_{N+M} (\theta^M / M!), \\ &\equiv \sum_{M=0}^\infty \xi^{N+M} x^{N+M} (\theta^M / M!) \quad [\text{by (1.5)}], \\ &\equiv \sum_{M=0}^\infty (\xi x)^N \cdot (\xi x)^M (\theta^M / M!) \quad [\text{by (1.124)}], \\ &\equiv (\xi x)^N \cdot \sum_{M=0}^\infty (\xi x)^M (\theta^M / M!) \quad [\text{by (1.125)}], \\ &\equiv (\xi x)^N \cdot e^{\xi x\theta}; \end{aligned}$$

$$(2.4) \quad \partial_\theta^N e^{\xi x\theta} \equiv (\xi x)^N \cdot e^{\xi x\theta},$$

in complete formal analogy with derivatives of ordinary (scalar) exponential functions. From (2.3), (2.4),

$$(2.5) \quad \partial_\theta^N [e^{\xi x\theta}]^A \equiv (A \cdot \xi x)^N e^{(A \cdot \xi x)\theta};$$

and from (1.101), (1.120),

$$(2.6) \quad e^{\xi x\theta} [e^{a\alpha\theta} + \cdots + e^{\gamma c\theta}] \equiv e^{(\xi x + \{a\alpha + \cdots + \gamma c\})\theta}.$$

The coefficient of $\theta^N / N!$ in the MacLaurin expansion of the left of (2.6) is in fact

$$\begin{aligned}
 & \sum_{s=0}^N \binom{N}{s} \xi^{N-s} x_{N-s} [\alpha^s a_s + \cdots + \gamma^s c_s], \\
 &= \sum_{s=0}^N \binom{N}{s} (\xi x)^{N-s} \{\alpha a + \cdots + \gamma c\}^s, \\
 &= (\xi x + \{\alpha a + \cdots + \gamma c\})^N,
 \end{aligned}$$

which is the coefficient of $\theta^N/N!$ on the right of (2.6).

Many of the more interesting applications to special sequences of numbers (like the Bernoulli or Euler numbers), arise in the following simple way. Let

$$\frac{\Lambda(\theta, \alpha, \cdots, \gamma)}{\Phi(\theta, \alpha, \cdots, \gamma)}$$

be a rational function of $\theta, \alpha, \cdots, \gamma$ in its lowest terms. Replace α, \cdots, γ by $\alpha e^{a\theta}, \cdots, \gamma e^{\gamma\theta}$, and let the MacLaurin expansion of the result be

$$\frac{\Lambda(\theta, \alpha e^{a\theta}, \cdots, \gamma e^{\gamma\theta})}{\Phi(\theta, \alpha e^{a\theta}, \cdots, \gamma e^{\gamma\theta})} \equiv \xi e^{x\theta},$$

thus defining the numbers x_N ($N = 0, 1, \cdots$). Let the MacLaurin expansions of Λ, Φ be

$$\Lambda(\theta, \alpha e^{a\theta}, \cdots, \gamma e^{\gamma\theta}) \equiv \eta e^{y\theta}, \quad \Phi(\theta, \alpha e^{a\theta}, \cdots, \gamma e^{\gamma\theta}) \equiv \xi e^{u\theta},$$

thus defining y_N, u_N . Hence

$$\begin{aligned}
 \eta e^{y\theta} &= \xi \xi e^{(x+u)\theta}, \\
 \eta y^N &= \xi \xi (x+u)^N.
 \end{aligned}$$

Hence, if $F(\theta)$ is a polynomial in θ , or a power series, if convergent,

$$F(\theta + y) = \xi \xi F(\theta + x + u),$$

in which, after expansion, exponents of y, x, u are degraded to suffixes.

In practice, the special notations $\{ \}, \dagger, () \cdot ()$, $A \cdot \alpha x$ are dropped, $+$, $() ()$, $A \alpha x$ being written, as the notation is a sufficient guide to the correct use of the algorithms. There are many extensions, in particular one to multiple suffixes, as in $x_{A,B}, \dots, c$, and the corresponding power series,

$$\sum_{A,B,\dots,C} x_{A,B,\dots,C} \alpha^A \beta^B \cdots \gamma^C.$$

Finally, everything down to (2.6) goes through unchanged if scalars in (1.1) are re-defined to be elements of any commutative ring with a modulus (\equiv identity with respect to multiplication).

THE ABELIAN QUASI-GROUP.*

By HARRIET GRIFFIN.

Introduction. The purpose of this paper is to investigate the abelian quasi-group, which is a commutative system of elements closed under a single operation, when certain conditions with respect to coset expansions are imposed by stated associative laws as explained in the first section. Throughout the paper we show how the abelian quasi-group differs from the abelian group. In section two we study in particular the minimal quasi-group of units both when each element is the unit for an element of the minimal quasi-group and also when this is not the case. In sections three and four we study the orders of elements in the case in which the minimal quasi-group is the identity element, and develop a method for setting up a quasi-group with an identity element and no subquasi-group other than the identity. We show that two conformal abelian quasi-groups need not be isomorphic.

The subquasi-groups of an abelian quasi-group under the conditions imposed form a Dedekind structure only in a special case as shown in section five and thus the abelian quasi-group differs greatly from the abelian group. Finally in section six we determine a necessary and sufficient condition that the cosets of an abelian quasi-group under the imposed associative laws shall form a quotient quasi-group.

SECTION I.

Associative laws of the abelian quasi-group.

To facilitate the reading of this paper, we begin with a connected account of certain fundamental properties of quasi-groups drawn largely from the paper of Hausmann and Ore¹ upon which our work is based. We do not always follow their exact wording, but we believe that any essential departure from their presentation is clearly indicated.

1. Fundamental definitions and notions of the finite abelian quasi-group Q . A groupoid is a system consisting of a set of distinct elements a, b, \dots and one binary operation (multiplication) such that to every ordered

* Received September 1, 1939; Revised November 20, 1939.

¹ B. A. Hausmann and Oystein Ore, "Theory of quasi-groups," *American Journal of Mathematics*, vol. 59 no. 4. (October, 1937),

pair of elements a, b , there corresponds a unique third (the product), $c = ab$, of the set.

If, further, to each ordered pair a, b there corresponds a unique x such that $ax = b$, and a unique y such that $ya = b$, the groupoid is called a quasi-group.

Since we are here interested in the abelian quasi-group, we impose the added condition that $ab = ba$. Then the quotients x and y above are equal.

No identity element need exist. However for each element a there is a unique e_a such that $ae_a = a$, called the unit for a .

If a is an element of Q , we define the powers of a as $a^n = a$ for $n = 1$ and $a^n = a^{n-1} \cdot a$ for $n > 1$. The order of a is then the least positive integer n for which $a^n = e_a$. Such a finite power of a exists, since Q is assumed finite.

Let a, b, \dots, k be any subset of Q . Then there is a least subquasi-group of Q which contains the elements a, b, \dots, k . We denote this subquasi-group by $\{a, b, \dots, k\}$. In particular the quasi-group $\{a\}$ generated by a single element is called a cyclic quasi-group. It is to be noted that $\{a\}$ may contain elements other than the powers of a .

2. Fundamental properties of Q . We wish the abelian quasi-group Q to have certain properties and hence impose the following conditions.

The expansion of Q by means of disjoint cosets with respect to any subquasi-group A is to exist, i. e., $Q = A + q_1A + \dots + q_nA$, where the q_i are in Q but not in A . The associative law expressing a necessary and sufficient condition for this property as proved by Hausmann and Ore is:

P_0 . If a and b are any elements of Q and c_0 and d_0 are determined so that

$$(ab)c_0 = ad_0,$$

then for any c

$$(ab)c = ad,$$

where d is an element of $\{c_0, d_0, c\}$.

Any element of a coset is to define the same coset. Again a necessary and sufficient condition for this property as proved by Hausmann and Ore is:

P_1 . For any elements a and b of Q

$$(ab)c = ad,$$

where d belongs to $\{b, c\}$.

It then follows from P_1 that $c(bC) = (bC)c = bC = (cb)C$ for any C containing c , and consequently each subquasi-group C of Q contains all the units of Q and each coset aC contains its multiplier a . Hence there is a subquasi-group of Q contained in all subquasi-groups of Q and containing all the units of Q . We call it the minimal subquasi-group E of Q . When Q con-

tains an element a such that $a^2 = a$, the minimal subquasi-group consists of a which is then the identity element. It is to be noted that the minimal quasi-group E is generated by any one of its elements, but it is not necessary that each of its elements be a unit.

Furthermore it is important to notice that P_1 implies P_0 .

The decomposition of Q into cosets is to be transitive, and herein we depart from the interpretation of transitivity given by Hausmann and Ore. By transitivity we mean that for every A and B such that $Q > A > B$, $Q = A + q_1A + \cdots + q_nA$, and $A = B + a_1B + \cdots + a_rB$, the expansion of Q/B can be obtained by substituting the expansion of A/B in the expansion of Q/A . A necessary and sufficient condition for this property is:

P_2 . When q is an element of $Q > \{a, B\}$ but is not in $\{a, B\}$, then for b in B ,

$$q(ab) = (qa)b_1,$$

where b_1 is generated by b .

Proof: (a). P_2 is a necessary condition.

Let

$$A = B + a_1B + \cdots + a_rB,$$

and

$$Q = A + q_1A + \cdots + q_nA,$$

where $Q > A > B$.

Then for transitivity of coset decomposition

$$Q = B + a_1B + \cdots + a_rB + q_1B + q_1(a_1B) + \cdots + q_n(a_rB).$$

If e is the unit for a_j , $q_i(a_j e) = q_i a_j$ and is in $q_i(a_j B)$. But since a coset is generated by any one of its elements,

$$q_i(a_j B) = (q_i a_j)B.$$

However for b in B , $\{b\} \leq B$. Hence, as for B , $q_i(a_j \{b\}) = (q_i a_j)(\{b\})$; so that for q in Q and not in $\{a, B\}$, $q(ab) = (qa)b_1$ where b generates b_1 .

(b). P_2 is a sufficient condition since as b varies over the elements of any B , b_1 varies over B , and thus $q(aB) = (qa)B$.

It is to be noted that P_2 does not govern all the products of elements of a quasi-group. Transitivity of coset expansion is not different from coset expansion except when applied to the product of elements of a proper subquasi-group by an element not in that subquasi-group. This fact is exhibited by Table X where having determined subquasi-group E as 1, 2, 3, and A as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, P_2 governs blocks of products like 13(4E), but it does not govern 7(4E) nor any of the products of 4 through 12 by any

of 4 through 12. This last set of elements forms what we shall call a free square in the multiplication table because the products are not governed by P_2 .

TABLE X. ABELIAN QUASI-GROUP OBEYING P_1 AND P_2 .

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	2	1	3	4	6	5	7	8	9	11	12	10	13	14	15	16	17	18	19	20	21	23	24	22
2	1	3	2	5	4	6	8	9	7	10	11	12	14	15	13	18	16	17	20	21	19	24	22	23
3	3	2	1	6	5	4	9	7	8	12	10	11	15	13	14	17	18	16	21	19	20	22	23	24
4	4	5	6	10	1	8	2	11	12	7	9	3	19	16	22	20	13	23	17	24	18	21	14	15
5	6	4	5	1	8	10	11	12	2	9	3	7	20	17	23	21	14	24	18	22	16	19	15	13
6	5	6	4	8	10	1	12	2	11	3	7	9	21	18	24	19	15	22	16	23	17	20	13	14
7	7	8	9	2	11	12	3	4	10	5	6	1	16	22	20	13	23	19	24	18	15	14	17	21
8	8	9	7	11	12	2	4	10	3	6	1	5	17	23	21	14	24	20	22	16	13	15	18	19
9	9	7	8	12	2	11	10	3	6	1	5	4	18	24	19	15	22	21	23	17	14	13	16	20
10	11	10	12	7	9	3	5	6	1	2	4	8	22	20	16	23	19	13	15	14	24	17	21	18
11	12	11	10	9	3	7	6	1	5	4	8	2	23	21	17	24	20	14	13	15	22	18	19	16
12	10	12	11	3	7	9	1	5	4	8	2	6	24	19	18	22	21	15	14	13	23	16	20	17
13	13	14	15	19	20	21	16	17	18	22	23	24	7	8	9	1	3	2	10	11	12	4	5	6
14	14	15	13	16	17	18	22	23	24	20	21	19	8	9	7	2	1	3	12	10	11	6	4	5
15	15	13	14	22	23	24	20	21	19	16	17	18	9	7	8	3	2	1	11	12	10	5	6	4
16	16	18	17	20	21	19	13	14	15	23	24	22	1	2	3	8	9	7	6	4	5	10	11	12
17	17	16	18	13	14	15	23	24	22	19	20	21	3	1	2	9	7	8	5	6	4	12	10	11
18	18	17	16	23	24	22	19	20	21	13	14	15	2	3	1	7	8	9	4	5	6	11	12	10
19	19	20	21	17	18	16	24	22	23	15	13	14	10	12	11	6	5	4	7	8	9	2	3	1
20	20	21	19	24	22	23	18	16	17	14	15	13	11	10	12	4	6	5	8	9	7	1	2	3
21	21	19	20	18	16	17	15	13	14	24	22	23	12	11	10	5	4	6	9	7	8	3	1	2
22	23	24	22	21	19	20	14	15	13	17	18	16	4	6	5	10	12	11	2	1	3	8	9	7
23	24	22	23	14	15	13	17	18	16	21	19	20	5	4	6	11	10	12	3	2	1	9	7	8
24	22	23	24	15	13	14	21	19	20	18	16	17	6	5	4	12	11	10	1	3	2	7	8	9

SECTION II.

The subquasi-group of units.

In this section we shall omit all proofs which refer to special tables. They may be found in the complete paper.

1. **The Cayley square.** It is to be noted that due to the symmetry of the Cayley square of an abelian quasi-group, if an element is not placed in the principal diagonal space of a row, the placing of the element causes a second row to be supplied with the element. On the other hand an element placed in the principal diagonal space of a row eliminates just that row. Hence if the order of the quasi-group, which is the number of elements it possesses, is even, an element which appears in the principal diagonal appears

there an even number of times. If the order of the quasi-group is odd, however, each element appears just once in the principal diagonal.

2. The quasi-groups \mathcal{E}_n . We consider first the minimal abelian quasi-group consisting of more than one element, i. e., where there is no element such that $a^2 = a$, and such that each element is the unit for one of the elements of the minimal quasi-group. We denote this quasi-group by \mathcal{E}_n , while the general minimal quasi-group is denoted by E or E_n where the subscript gives the order.

Since any element has but a single unit, and since in \mathcal{E}_n each element is the unit for one element of \mathcal{E}_n , when a is the unit for b , b must be the unit for some c , c for d , etc. If we continue in this fashion and include all the elements of \mathcal{E}_n upon returning to a , we shall say that the units set up a single cycle. Otherwise it is clear that the set of units will be separated into two or more cycles. It is evident, however, that no cycle may have but two elements since if a is the unit for b , $ab = b$, and then b cannot be the unit for a . Hence in the case of \mathcal{E}_3 there can be but a single cycle.

THEOREM 1. *There are no \mathcal{E}_n of order 2 nor 4.*

THEOREM 2. *An \mathcal{E}_n exists for every n not 2 nor 4.*

Proof. By first setting up the cycle of units, 2 for 1, 3 for 2, \dots , n for 1, and then the principal diagonal, we have developed general methods for building abelian quasi-groups of any odd order, of odd order greater than 5, and of even order greater than 6. \mathcal{E}_6 is a special case. There is no subquasi-group in these tables since the unit for an element must be in any subquasi-group with it.

THEOREM 3. *All \mathcal{E}_3 and all \mathcal{E}_5 are abstractly identical.*

Proof. Having chosen the units the tables are uniquely determined.

THEOREM 4. *The \mathcal{E}_6 are abstractly identical.*

Proof. The set of units must form a single cycle, and the two seemingly distinct quasi-groups that it is possible to build on the cycle of units are cyclic permutations of one another.

THEOREM 5. *Although the set of units for \mathcal{E}_7 must form a single cycle, the \mathcal{E}_7 are not abstractly identical.*

THEOREM 6. *The \mathcal{E}_n where $n > 7$ are not abstractly identical.*

Proof. For any even order we can set up an abelian quasi-group \mathcal{E}_n in which the set of units is broken into cycles 1 through $n-3$ and $n-2$

through n . For odd integers (not 3 nor 7) of the form $4n + 3$ a second method shows how to break the set of units of \mathcal{E}_{4n+3} into cycles 1 through $2n + 1$, and $2n + 2$ through $4n + 3$, while for odd integers greater than 5 of the form $4n + 1$ we can set up an \mathcal{E}_{4n+1} with the set of units broken into the cycles 1 through $2n + 2$, and $2n + 3$ through $4n + 1$. In all cases the diagonal elements are so chosen that there is no subquasi-group.

These methods together with Theorems 2 and 5 show that there are at least two distinct \mathcal{E}_n for any $n \geq 7$ since a quasi-group in which the set of units is broken into two cycles cannot be isomorphic with one in which a single cycle includes all elements.

3. The order of an element of the minimal quasi-group.

THEOREM 7. *No element of \mathcal{E}_n can be of order 1 nor n .*

Proof. The order of an element cannot be one since $a^2 \neq a$. It cannot be n since the element for which a is the unit cannot occur among the powers of a .

THEOREM 8. *In a minimal quasi-group E_n if an element e is the unit for just $s \geq 0$ elements of E , the order of e is at most $n - s$, and it cannot be $n - s - 1$.*

Proof. Let e be the unit for s distinct elements. Then since none of these s elements can be powers of e , the order of e is at most $n - s$, and may be $n - s$ if the powers of e exhaust the $n - s$ elements for which e is not the unit. If the order of e were $n - s - 1$, the remaining element when multiplied by e would have to give itself and e would be the unit for $s + 1$ elements. On the other hand if the order of e is less than $n - s - 1$, e need not be the unit for any of the remaining elements.

COROLLARY. *In E_n an element cannot be the unit for $n - 1$ elements.*

When n is 6 or 7, we have set up E_n in which there are elements which are units for one or two elements and which are of all orders not excluded by Theorem 8. We also have examples of elements which are not units for any element of E_n such that their orders include all integers from 2 through n except $n - 1$.

SECTION III.

The abelian quasi-group with an identity element and no subquasi-group other than the identity.

1. Order of I_n . If an abelian quasi-group of order n has no subquasi-group except the identity element 1, we designate it by I_n .

THEOREM 1. *There are no abelian quasi-groups I_n of orders 2, 3, nor 5 except when they are groups, and there are none of even order > 2 .*

THEOREM 2. *An I_n exists for every odd order > 5 .*

Proof. Set up the multiplication table of a cyclic group by putting 1 through n in the first column and row and by filling in all diagonals from the upper right to the lower left with the number in the first row of the diagonal. Make the last column $n, 1, 2, 3, \dots, n-1$, and carry the diagonals through as before. Then by the method of formation the powers of 2 are $2, 3, 4, \dots, n, 1$, so that 2 is always of order n . Now interchange the elements of the last row with the principal diagonal elements immediately above them. This operation makes n the identity, so call 1, n and $n, 1$ throughout the table. The order of 2 remains n . The result is quasi-group L .

There is no subquasi-group other than 1 since for any element except 1 and 2, $a^2 = a - 1$, and hence any element not 1 generates 2. But 2 is of the n -th order. Therefore each element except 1 generates the quasi-group L .

Furthermore the quasi-group is not a group since $2(3 \cdot 4) = 1$, while $(2 \cdot 3)4 = 3$ and this part of the table remains for all orders greater than 5.

2. The order of elements of L . The manner of forming L makes it possible to find the orders of its elements.

THEOREM 3. *The order of every element a except 1, 2, and n of the quasi-group L of order n set up by Theorem 2 is n or $n/g + 1$, where g is the greatest common divisor of $a - 1$ and n , according as $a - 1$ is or is not prime to n .*

THEOREM 4. *The order of the element n of quasi-group L is less than n except when n is a prime and 2 is a primitive root mod n in which case the order of n is n .*

3. The order of the elements of I_n .

THEOREM 5. *There is no element of I_n of order 2 nor $n - 1$.*

In some cases it is possible to set up a quasi-group lacking a subquasi-group except 1 without interchanging all the elements of the principal diagonal and the last row as in Theorem 2. In particular except when n is a prime and 2 is a primitive root mod n , there is a set of less than n interchanges which include the 1 and n which always gives a quasi-group. We have used these methods to set up I_n when n is 7, 9, and 15 and these quasi-groups afford examples of elements of every order except 2 and $n - 1$.

SECTION IV.

The abelian quasi-group in general.

We turn now to some facts about the general abelian quasi-group Q_n .

THEOREM 1. *If a quasi-group Q_n has an element of order $n - 1$, the element generates the quasi-group and there is no identity element.*

THEOREM 2. *Two quasi-groups of the same order and having elements of corresponding orders need not be abstractly identical.*

Proof. We demonstrate this fact by an illustration. Take R with the elements 1, 2, 3, 4, 5, 6 such that the products of these elements in order by 1 are 5, 2, 3, 1, 6, 4; by 2 are 2, 6, 1, 4, 5, 3; by 3 are 3, 1, 4, 5, 2, 6; by 4 are 1, 4, 5, 6, 3, 2; by 5 are 6, 5, 2, 3, 4, 1; and by 6 are 4, 3, 6, 2, 1, 5. Take S with elements 2, 3, and 4 as above but let the products for 1 be 6, 2, 3, 1, 4, 5; for 5 be 4, 5, 2, 3, 6, 1; and for 6 be 5, 3, 6, 2, 1, 4. Since 3 is in each the only element of order five, if the quasi-groups were isomorphic, these elements would have to correspond to each other. But then 4 of R must correspond to 4 of S ; 5 to 5; 2 to 2; and 1 to 1. But in R $1 \cdot 1 = 5$, while in S $1 \cdot 1 = 6$.

This fact is interesting since we recall that two abelian groups which are conformal are isomorphic.

Under certain conditions which we discuss in Section VI the cosets with respect to a subquasi-group B of Q always appear in blocks throughout the multiplication table due to the fact that Q/B forms a quotient quasi-group. In this case the cosets themselves form an abelian quasi-group with an identity element consisting of the subquasi-group B , and since the blocks must combine as do any elements of the blocks, the blocks obey the associative law of the original elements. We can under these circumstances make certain general statements about the order of elements of Q .

THEOREM 3. *The order of any element of Q_n with blocks of cosets with respect to a subquasi-group B_m throughout the table may be only:*

a. *Those orders permitted in the identity element $I = B$ of the quotient quasi-group Q/B .*

b. *1, 2, \dots , or m times the order of any block except I of Q/B .*

If the cosets themselves form a quasi-group without a subquasi-group except I , the theorems of Section III apply. But what is the order of a coset of the quotient quasi-group if there are subquasi-groups other than I ? This

question is the same as asking: What is the order of an element of a quasi-group of order n with an identity element where there may be subquasi-groups other than 1, and where the blocks of cosets with respect to any subquasi-group need not appear throughout the Cayley square? When there is an identity element, beyond the fact that the order of an element cannot be $n - 1$, and that if n is odd, there can be no element of order 2 (due to the coset expansions), we can state no law which the order of an element obeys. On the other hand if Q_n has a minimal subquasi-group of units $E < Q_n$ and the block formation does not prevail throughout the multiplication table, no elements of E are of order $n - 1$, and the order of an element a not in E cannot be $n - 1$, since if b were the element not among the $n - 1$ powers of a , $ab = b$, and then a would be a unit.

Herein lies the difference between the multiplication table of the abelian group and the abelian quasi-group, for the cosets of an abelian group always form a quotient group and hence the elements of the cosets form blocks throughout the multiplication table. But more than that, due to the associative law $a(bc) = (ab)c$, the elements take the same positions in the blocks throughout the table. It is this orderly characteristic of the Cayley square for the abelian group which distinguishes it from that of the abelian quasi-group.

SECTION V.

Structure theory.

1. Structure definition is satisfied. A structure is a partially ordered system in which every two elements have a union and a cross cut. In the case of the quasi-group Q the union $[A_1, \dots, A_n]$ of subquasi-groups of Q is the smallest subquasi-group which contains each A_i , while the cross cut (A_1, \dots, A_n) is the largest subquasi-group contained in every A_i . It is evident that the subquasi-groups of an abelian quasi-group form a finite structure.

2. Dedekind structure. In order that a structure be a Dedekind structure, it must satisfy the axiom:

When A, B, C are any three elements of a structure such that $A < C < [A, B]$, then

$$C = [A, (B, C)].$$

When the abelian quasi-group has the properties given by P_1 and P_2 , the subquasi-groups do not in general satisfy this axiom. To show that it is violated we use the abelian quasi-group W of twenty-four elements built as the direct product of two of its subquasi-groups. We let A be the subquasi-group of elements 1, 2, 3, 4, 5, 6; B be 1, 2, 3, 13, 14, 15; C be 1, 2, 3, 4, 5, 6, 7,

8, 9, 10, 11, 12; while the minimal subquasi-group of units is 1, 2, 3. Although $[A, B] = W$, $[A, (B, C)] = A \neq C$. Hence the subquasi-groups of W , which obeys the laws of coset expansion and transitivity of coset decomposition, do not in general form a Dedekind structure.

The simple case where every subquasi-group of Q belongs to the single principal chain $Q > A_1 > A_2 \cdots > E$ is an exception since in this case every element of $[A_1, A_2]$ is an a_1 and the Dedekind axiom must be satisfied by the method of the following Theorem 2.

In view of the fact that the subgroups of an abelian group always form a Dedekind structure and since coset expansions and transitivity of coset decomposition are such outstanding properties of the group, it is interesting to find that the subquasi-groups of an abelian quasi-group having these properties fail to form necessarily a Dedekind structure.

If, however, we strengthen P_2 to read:

P_3 . For any three elements a, b, c of Q

$$a(bc) = (ab)c_1,$$

where c_1 is an element of $\{c\}$, we have, as proved by Hausmann and Ore, a Dedekind structure.

THEOREM 1. When P_3 holds, the elements of $[A, B]$ take the form ab .

Proof. Any $a = ae_a$. Furthermore:

$$(a_1b_1)(a_2b_2) = ((a_1b_1)a_2)b_3 = ((a_1a_2)b_4)b_3 = a_3(b_4b_3) = a_3b_5$$

where b_3 is generated by b_2 ; b_4 by b_1 ; etc.

It follows then as proved by Hausmann and Ore that:

THEOREM 2. When $A > B$, the abelian subquasi-groups A, B, C of Q obeying P_3 satisfy the Dedekind relation

$$(A, [B, C]) = [B, (A, C)].$$

Therefore, as pointed out by Hausmann and Ore, the analogues of the Zassenhaus-Schreier refinement theorem, of the Jordan-Holder theorem on the invariance of the lengths of principal chains, and of the Schmidt-Remak theorem on direct decomposition must hold for the subquasi-groups of an abelian quasi-group obeying P_3 . We can say, as does Ore with respect to the group, that in these respects the theory of the abelian quasi-group is more a property of the subquasi-groups than of the elements themselves.

SECTION VI.

The quotient quasi-group.

1. The abelian quotient quasi-group Q/B . When all and only the elements of a subquasi-group B of Q occupy the upper left-hand corner of the Cayley square of Q , due to P_1 there are always blocks of cosets of B at the top and left side of the table of Q . If these blocks are maintained throughout the table, they form an abelian quotient quasi-group with the identity element B . In order that blocks of cosets with respect to B be maintained throughout the table of Q any element of the coset q_1B when multiplied by an element of q_2B must give another coset and that coset must be the one that contains q_1q_2 , which is $(q_1q_2)B$. Conversely if $(q_1B)(q_2B) = (q_1q_2)B$, there are blocks of cosets with respect to B throughout the table of Q . Hence $(q_1B)(q_2B) = (q_1q_2)B$ is a necessary and sufficient condition for blocks of cosets with respect to the subquasi-group B throughout the multiplication table of Q which obeys P_1 and P_2 .

2. The quotient quasi-group under P_3 . Under the law P_3 we have:

$$(aC)(bC) = ((aC)b)C = ((ab)C)C = (ab)C.$$

Furthermore if $(aC)(bC) = (ab)C$, by applying P_3 we have:

$$(ab)C = ((bC)a)C.$$

But $a(bC)$ is a coset with respect to C , and if it is multiplied by an element of C , it must give one of its own set. Therefore

$$(ab)C = a(bC).$$

Thus if $C = \{c\}$, for any a, b, c , $(ab)c = a(bc_1)$, where c generates c_1 .

In like manner if we assume for any a, b, c , that $(ab)c = a(bc_1)$ where c_1 is generated by c , it follows that:

$$(ab)C = ((ab)C)C = ((bC)a)C = (bC)(aC).^2$$

Furthermore

$$a(bC) = (a(bC))C = (bC)(aC) = (ab)C.$$

Hence it follows that:

THEOREM 1. *If for every a, b, c , $(ab)c = a(bc_1)$ where c_1 is generated by c , then there is a quotient quasi-group with respect to any subquasi-group C and further c is generated by c_1 .*

² Hausmann and Ore, *loc. cit.*

3. The quotient quasi-group under P_2 . If the abelian quasi-group obeys the law P_2 , we do not in general have blocks of cosets throughout the table since there are the free squares which are not governed by P_2 . However it is interesting to notice that P_2 forces products of an element not in a subquasi-group B of Q by the elements of a coset of B/E where E is the minimal subquasi-group to give the elements of a coset of B/E .

Consider subquasi-groups B and E of Q such that $E < B < Q$. Then P_2 applies to the product of q not in B by any two elements of B and $q(bc) = (qb)c_1$. When all the cosets of Q/E have been determined, the product qb is determined only in the case where b is an e . Take $b \neq e$ of the coset bE . Let $qb = d$ which may not be in qE nor in bE . Then $q(be_b) = d = (qb)e_{qb}$. For another e , $q(be_s) = (qb)e_r = de_r$, and $e_r \neq e_{qb}$. Hence as e_s varies through the e , so does e_r , and de_r must vary through the elements of dE . Hence the products of q by bE give another coset dE . These cosets must then combine to form cosets with respect to the next larger subquasi-group.

It is to be noted, however, that another element of qE when multiplied by bE may give the elements of a coset different from dE . This fact is exhibited by quasi-group X , and shows that blocks of cosets with respect to E need not exist in this part of the multiplication table.

Let C be a cyclic subquasi-group next larger than E . Then all elements of C not in E generate C . When we assume $q(bc) = (qb)c_1$ where q is not in $\{b, C\}$, if c is in E , then c_1 is in E and one generates the other. But if c is in C and not in E , then too c_1 must generate c , and as c varies through these c 's, c_1 must vary through the same c 's. After repeated applications of this argument we may conclude:

THEOREM 2. *For q not in $\{b, C\}$ and c in C , $q(bc) = (qb)c_1$ where c generates c_1 implies that c_1 generates c , and where c_1 generates c , c generates c_1 .*

If we assume P_2 , how may we strengthen this law in order to have a quotient quasi-group with respect to any subquasi-group of Q ? First consider the case above where $qb = d$ and which showed that the elements $q(bE)$ form a coset with respect to E but that blocks of cosets are not necessarily formed. We may take these as the elements of a row. To complete the block we consider that $qb = bq$ and $b(qe_q) = (bq)e_{bq}$. Then if we change e_q , qe_r varies through the coset qE and the products are to give dE so that $b(qe_r)$ must give $(bq)e_s$ where e_s varies through all the e 's. Hence if the column is to be dE , $b(qe) = (bq)e_1$.

If we consider the blocks which are to be obtained by multiplying q_1 by qE where neither q_1 nor q is in B , a like argument permits us to conclude that if we have a quotient quasi-group with respect to E , then $a(be) = (ab)e_1$ for any a, b, e .

In order to obtain a condition for a quotient quasi-group with respect to every subquasi-group, we take $C > E$ with no subquasi-group between C and E . Then as noted, any c in C but not in E generates C . We assume that we have blocks with respect to E . Then $(qC)(q_1C)$, where q and q_1 are not in C , must give $(qq_1)C$, i. e., $q(q_1C)$ must equal $(qq_1)C$. But $q(q_1c) = (qq_1)c_1$. Therefore $q(q_1c) = (qq_1)c_1$ where c generates c_1 , for any c of C but not in E . If we call these elements a set of a row, $q_1(qC)$ will fill the column and give again c_1 generated by c . By repeating this process we see that:

THEOREM 3. *When an abelian quasi-group satisfies P_1 and P_2 , if a quotient quasi-group exists for every C , then for every a, b, c ,*

$$a(bc) = (ab)c_1,$$

where c generates c_1 .

Together with Theorem 1 of this section we now have:

THEOREM 4. *When an abelian quasi-group satisfies P_1 and P_2 , a necessary and sufficient condition for a quotient quasi-group with respect to every subquasi-group C of Q is:*

For any a, b, c ,

$$a(bc) = (ab)c_1,$$

where c generates c_1 .

NEW YORK UNIVERSITY.

REFERENCES.

-
- Oystein Ore, "On the foundation of abstract algebra," I, *Annals of Mathematics*, vol. 36 (1935), pp. 406-437; II, *ibid.*, vol. 37 (1936), pp. 265-292; "Structures and group theory," *Duke Mathematical Journal*, vol. 3 (1937), pp. 149-174; "On the application of structure theory to groups," *Bulletin of the American Mathematical Society*, vol. 44 (1938), pp. 801-806.
- Hausmann and Ore, "Theory of quasi-groups," *American Journal of Mathematics*, vol. 59 (1937), pp. 983-1004.

THE GAUSSIAN LAW OF ERRORS IN THE THEORY OF ADDITIVE NUMBER THEORETIC FUNCTIONS.*¹

By P. ERDÖS and M. KAC.

The present paper concerns itself with the applications of statistical methods to some number-theoretic problems. Recent investigations of Erdős and Wintner² have shown the importance of the notion of statistical independence in number theory; the purpose of this paper is to emphasize this fact once again.

It may be mentioned here that we get as a particular case of our main theorem the following result:

If $\nu(m)$ denotes the number of prime divisors of m , and K_n the number of those integers from 1 up to n for which $\nu(m) < \lg \lg n + \omega \sqrt{2 \lg \lg n}$ (ω an arbitrary real number), then

$$\lim_{n \rightarrow \infty} \frac{K_n}{n} = \pi^{-\frac{1}{2}} \int_{-\infty}^{\omega} \exp(-u^2) du.$$

This theorem refines some known results of Hardy, Ramanujan³ and Erdős.⁴

1. In what follows p will denote a prime and ω will denote a real number.

Let $f(m)$ be an additive number-theoretic function, so that $f(mn) = f(m) + f(n)$ if $(m, n) = 1$. Suppose that $f(p^a) = f(p)$ and $|f(p)| \leq 1$. Obviously

$$f(m) = \sum_{p|m} f(p).$$

Furthermore put $\sum_{p < n} p^{-1} f(p) = A_n$ and $(\sum_{p < n} p^{-1} f^2(p))^{1/2} = B_n$. Then our main theorem may be stated as follows:

* Received December 7, 1939.

¹ A preliminary account appeared in the *Proceedings of the National Academy*, vol. 25 (1939), pp. 206-207.

² P. Erdős and A. Wintner, "Additive arithmetic functions and statistical independence," *American Journal of Mathematics*, vol. 61 (1939), pp. 713-722.

³ Srinivasa Ramanujan, *Collected Papers* (1927), pp. 262-275.

⁴ P. Erdős, "Note on the number of prime divisors of integers," *Journal of the London Mathematical Society*, vol. 12 (1937), pp. 308-314.

THEOREM. If $B_n \rightarrow \infty$ as $n \rightarrow \infty$, and K_n denotes the number of integers m from 1 up to n for which

$$f(m) < A_n + \omega \sqrt{2} B_n$$

then

$$\lim_{n \rightarrow \infty} \frac{K_n}{n} = \pi^{-1/2} \int_{-\infty}^{\omega} \exp(-u^2) du = D(\omega).$$

2. We first prove the following

LEMMA 1. Let

$$f_l(m) = \sum_{\substack{p|m \\ p < l}} f(p).$$

Then denoting by δ_l the density of the set of integers m for which $f_l(m) < A_l + \omega \sqrt{2} B_l$ one has

$$\lim_{l \rightarrow \infty} \delta_l = D(\omega).$$

Let $\rho_p(n)$ be 0 or $f(p)$ according as p does not or does divide n . Then

$$f_l(m) = \sum_{p < l} \rho_p(m).$$

Since the $\rho_p(n)$ are statistically independent, $f_l(m)$ behaves like a sum of independent random variables and consequently the distribution function of $f_l(m) - A_l/\sqrt{2} B_l$ is a convolution (Faltung) of the distribution functions of $\rho_p(m) - p^{-1}f(p)/\sqrt{2} B_l$ ($p < l$). It is easy to see that the "central limit theorem of the calculus of probability" can be applied to the present case,⁵ and this proves our lemma.

3. Lemma 1 is the only "statistical" lemma in the proof. Using this lemma, the main result will be established by purely number-theoretical methods.

LEMMA 2. If m_n tends to ∞ (as $n \rightarrow \infty$) more rapidly than any fixed

⁵ Loc. cit. 2, where statistical independence of arithmetical functions is defined and discussed. See also P. Hartman, E. R. van Kampen and A. Wintner, *American Journal of Mathematics*, vol. 61 (1939), pp. 477-486.

⁶ Cf. for instance the first chapter of S. Bernstein's paper, "Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes," *Mathematische Annalen*, vol. 97, pp. 1-59. See also M. Kac and H. Steinhaus, "Sur les fonctions indépendantes II," *Studia Math.*, vol. 6 (1936), pp. 59-66.

power of s_n , then the number of integers from 1 up to m_n which are not divisible by any prime less than s_n is equal to

$$\frac{m_n e^{-C}}{\lg s_n} + o\left(\frac{m_n}{\lg s_n}\right),$$

where C denotes Euler's constant.

The proof of this statement is implicitly contained in the reasoning of V. Brun on page 21 of his famous memoir "Le crible d'Erasosthène et le théorème de Goldbach"⁷ and may therefore be omitted.

Let $\phi(n)$ represent a function which tends, as $n \rightarrow \infty$, to 0 in such a way that $n^{\phi(n)} \rightarrow \infty$. The function $n^{\phi(n)}$ will be denoted by α_n and $n^{\sqrt{\phi(n)}}$ by β_n . Let $a_1(n), a_2(n), \dots$ be the integers whose prime factors are all less than α_n , and let $\psi(m; n)$ be the greatest a_i which divides m . We then have the following

LEMMA 3. The number of integers $m \leq n$ for which $\psi(m; n) = a_i(n)$, where $a_i(n) \leq \beta_n$ is equal to

$$\frac{e^{-Cn}}{a_i(n)\phi(n)\lg n} + o\left(\frac{n}{a_i(n)\phi(n)\lg n}\right). \quad \text{unf. in } i.$$

This is a direct consequence of Lemma 2. For consider all those integers $\leq n$ which are of the form $r \cdot a_i(n)$ and such that r is not divisible by any prime $< \alpha_n$. Evidently, the integers thus defined are all the integers $\leq n$ for which $\psi(m; n) = a_i(n)$. Their number is equal to the number of integers r which are $\leq n/a_i(n)$ and not divisible by any prime $< \alpha_n$. The restriction $a_i(n) < \beta_n$ makes $n/a_i(n)$ tend to ∞ more rapidly than any power of α_n and therefore Lemma 2 can be applied (put $m_n = n/a_i(n)$ and $s_n = \alpha_n$). This completes the proof.

LEMMA 4. The number y of integers $\leq M$ divisible by an $a_i(n) > \beta_n$ is less than $bM\sqrt{\phi(n)}$, where b is an absolute constant. (It follows from this that the density of the integers which are divisible by an $a_i(n) > \beta_n$ is less than $b\sqrt{\phi(n)}$.)

We have

$$\prod_{m=1}^M \psi(m; n) = \prod_{p < \alpha_n} p \sum_{r=1}^{\infty} [M/p^r] < \prod_{p < \alpha_n} p^{2M/p};$$

and since

$$\lg \prod_{p < \alpha_n} p^{2M/p} = 2M \sum_{p < \alpha_n} p^{-1} \lg p \sim 2M\phi(n) \lg n$$

⁷ Skrifter Videns, Kristiania, 1920.

one has

$$\prod_{m=1}^M \psi(m; n) < n^{bM\phi(n)}.$$

Hence, finally

$$(\beta_n)^y = (n^{\sqrt{\phi(n)}})^y < n^{bM\phi(n)}, \text{ i. e., } y < bM\sqrt{\phi(n)}.$$

4. LEMMA 5. Denote by l_n the number of integers from 1 up to n for which

$$(i) \quad f_{a_n}(m) < A_{a_n} + \omega\sqrt{2} B_{a_n}.$$

Then

$$\lim_{n \rightarrow \infty} \frac{l_n}{n} = D(\omega).$$

Divide the integers from 1 up to n which satisfy (i) into classes E_1, E_2, \dots so that m belongs to E_i if and only if $\psi(m; n) = a_i(n)$; and denote by $|E_i|$ the number of integers in E_i . One obviously has

$$l_n = \sum_i |E_i| = \sum_{a_i \leq \beta_n} |E_i| + \sum_{a_i > \beta_n} |E_i|.$$

By Lemma 4 $\sum_{a_i > \beta_n} |E_i| < bn\sqrt{\phi(n)}$ and therefore it is sufficient to prove that $n^{-1} \sum_{a_i \leq \beta_n} |E_i| \rightarrow D(\omega)$ as $n \rightarrow \infty$. On the other hand by Lemma 3

$$(ii) \quad \sum_{a_i \leq \beta_n} |E_i| = \left(\frac{e^{-C} \cdot n}{\phi(n) \lg n} + o\left(\frac{n}{\phi(n) \lg n}\right) \right) \sum'_{a_i \leq \beta_n} \frac{1}{a_i(n)},$$

where the dash in the summation indicates that it is extended over the a_i 's satisfying $f_{a_n}(a_i) < A_{a_n} + \omega\sqrt{2} B_{a_n}$. In order to evaluate \sum' , divide all the integers into classes F_1, F_2, \dots having the property that m belongs to F_i if and only if $\psi(m; n) = a_i(n)$ and let $\{F_i\}$ denote the density of F_i . Consider now the set $\sum' F_i$, where the dash in summation has the same meaning as above. By putting $l = \alpha_n$ and using Lemma 1 we have that $\{\sum' F_i\} \rightarrow D(\omega)$ as $n \rightarrow \infty$ or $\{\sum' F_i\} = D(\omega) + o(1)$. Now

$$(iii) \quad \sum'_{a_i \leq \beta_n} F_i = \sum'_{a_i \leq \beta_n} F_i + \sum'_{a_i > \beta_n} F_i$$

and by Lemma 4

$$(iv) \quad \left\{ \sum'_{a_i > \beta_n} F_i \right\} < b\sqrt{\phi(n)}.$$

Furthermore there is only a finite number of a_i 's which are less than β_n and therefore $\left\{ \sum'_{a_i < \beta_n} F_i \right\} = \sum'_{a_i < \beta_n} \{F_i\}$. But

$$\{F_i\} = \frac{1}{a_i(n)} \prod_{p < a_n} \left(1 - \frac{1}{p}\right) = \frac{1}{a_i(n)} \left(\frac{e^{-C}}{\phi(n) \lg n} + o\left(\frac{1}{\phi(n) \lg n}\right) \right)$$

and this implies that

$$(v) \quad \left\{ \sum'_{a_i < \beta_n} F_i \right\} = \left(\frac{e^{-C}}{\phi(n) \lg n} + o \left(\frac{1}{\phi(n) \lg n} \right) \right) \sum'_{a_i < \beta_n} \frac{1}{a_i(n)}.$$

Finally (iii), (iv) and (v) give

$$D(\omega) - b\sqrt{\phi(n)} < \left(\frac{e^{-C}}{\phi(n) \lg n} + o \left(\frac{1}{\phi(n) \lg n} \right) \right) \sum'_{a_i < \beta_n} \frac{1}{a_i(n)} < D(\omega) + \theta(1).$$

The combination of this formula with (ii) completes the proof of our Lemma.

5. We now come to the proof of the main theorem. Notice first that for $m \leq n$, $|f(m) - f_{a_n}(m)| < 1/\phi(n)$. In fact, $|f(p)| \leq 1$ implies that $|f(m) - f_{a_n}(m)|$ is less than the number of those prime divisors of m which are $\geq \alpha_n$. This number is obviously $< 1/\phi(n)$, since $(\alpha_n)^{1/\phi(n)} = n$. Notice furthermore that $|f(p)| \leq 1$ and the well known results concerning the sum $\sum_{p < n} p^{-1}$ imply that $|A_n - A_{a_n}| < -C_1 \lg \phi(n)$ and $|B_n - B_{a_n}| < -C_2 \lg \phi(n)$, where C_1 and C_2 are absolute constants.

Now choose $\phi(n)$ so that $1/\phi(n) = o(B_n)$. Evidently every $m \leq n$ satisfying the inequality $f(m) < A_n + \omega\sqrt{2} B_n$ also satisfies, for sufficiently large n , the inequality $f_{a_n}(m) < A_{a_n} + (\omega + \epsilon)\sqrt{2} B_{a_n}$. In addition every $m \leq n$ satisfying $f_{a_n}(m) < A_{a_n} + (\omega - \epsilon)\sqrt{2} B_{a_n}$ satisfies, for sufficiently large n , the inequality $f(m) < A_n + \omega\sqrt{2} B_n$. Hence, by Lemma 5,

$$D(\omega - \epsilon) \leq \liminf \frac{K_n}{n} \leq \limsup \frac{K_n}{n} \leq D(\omega + \epsilon);$$

and this proves the theorem, since $\epsilon > 0$ is arbitrary.

6. The theorem mentioned in the introduction is obviously a particular case of our main theorem. It corresponds to the case $f(p) = 1$. Because of the large number of applications of $\nu(m)$ it is of special interest. It should be mentioned that the assumption $f(p^a) = f(p)$ can be removed; also $|f(p)| \leq 1$ may be replaced by a much weaker condition. This however, would complicate the statement of the main theorem.

We may perhaps point out that Lemma 2 (Brun) is the "deepest" part of the proof and that the "statistical" part is relatively superficial. However, the statistical considerations seemed to be suggestive and fruitful in leading to new and perhaps striking results.

INSTITUTE FOR ADVANCED STUDY AND CORNELL UNIVERSITY.

ON THE STANDARD DEVIATIONS OF ADDITIVE ARITHMETICAL FUNCTIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

1). 1. If $p = p_k$ denotes the k -th prime number and l a positive integer, then, on starting with any double sequence of real numbers a_{lk} , put, for every positive integer n ,

$$(1) \quad f(n) = \sum_{k=1}^{\infty} f_k(n) = \lim_{k \rightarrow \infty} \tilde{f}_k(n),$$

where

$$(2) \quad \tilde{f}_k(n) = \sum_{j=1}^k f_j(n),$$

and

$$(3) \quad f_k(n) = \begin{cases} 0, & \text{if } n \not\equiv 0 \pmod{p_k}, \\ f(p_k^l), & \text{if } p_k^l | n \text{ and } p_k^{l+1} \nmid n, \end{cases}$$

finally $f(p_k^l) = a_{lk}$. It is clear that the functions $f(n)$, thus obtained, and only these functions, are additive, i. e., such that

$$(4) \quad f(n_1 n_2) = f(n_1) + f(n_2) \text{ whenever } (n_1, n_2) = 1; \quad (f(1) = 0).$$

In fact, the series (1) is convergent for every choice of the double sequence $\{a_{lk}\}$, since the series has, for every fixed n , at most a finite number of non-vanishing terms. It is also clear that an additive function f and either of the two sequences of additive functions $\{f_k\}$, $\{\tilde{f}_k\}$ of n determine each other uniquely. The additive functions to be considered will be assumed to be *real-valued*.

For a given $y = f(n)$, define $y^* = f^*(n)$ by placing

$$(5) \quad y^* = y \text{ or } y^* = 1 \text{ according as } |y| < 1 \text{ or } |y| \geq 1.$$

Then the question as to the existence of an asymptotic distribution function of an f may be answered as follows:¹

(I) An additive $f(n)$ has an asymptotic distribution function $\sigma(x)$, $-\infty < x < +\infty$, if and only if both series

$$(6_1) \quad \sum_p \frac{f^*(p)}{p};$$

$$(6_2) \quad \sum_p \frac{f^*(p)^2}{p}$$

are convergent.

* Received April 8, 1940.

¹ P. Erdős and A. Wintner, "Additive arithmetical functions and statistical independence," *American Journal of Mathematics*, vol. 61 (1939), pp. 713-721.

It is clear that if f_k is an additive function depending only on the k -th prime number (i. e., if it is of the type (3)), then it always has the asymptotic distribution function

$$(7) \quad \sigma_k(x) = \frac{p}{p-1} \sum_{f(p^m) < x} \frac{1}{p^m}; \quad -\infty < x < \infty, \quad (p = p_k; m = 0, 1, \dots).$$

Furthermore, it is easy to see that the terms $\tilde{f}_k(n)$ of an additive function (2), which depends on a finite number of prime numbers, are statistically independent; so that, in particular, the additive function \tilde{f}_k always has an asymptotic distribution function $\tilde{\sigma}_k(x)$, $-\infty < x < \infty$, and the latter is represented by the convolution

$$(8) \quad \tilde{\sigma}_k = \sigma_1 * \sigma_2 * \dots * \sigma_k.$$

It was shown *loc. cit.*¹ that the infinite sum (1) cannot have an asymptotic distribution function $\sigma(x)$ unless it has the asymptotic distribution which one would formally expect in virtue of (1) and (8), i. e., that (I) may be replaced by the following theorem:

(I') *An additive $f(n)$ has an asymptotic distribution function $\sigma(x)$, $-\infty < x < +\infty$, if and only if the infinite convolution $\sigma_1 * \sigma_2 * \dots$ of the asymptotic distribution functions (7) of its terms (3) is convergent, in which case one necessarily has*

$$(9) \quad \sigma = \sigma_1 * \sigma_2 * \dots.$$

2. For the more restricted class of almost periodic functions (B^2), the following theorem was recently established:²

(II) *An additive $f(n)$ is almost periodic (B^2) if and only if both series*

$$(10_1) \quad \sum_p \frac{f(p)}{p}; \quad (10_2) \quad \sum_{l=1}^{\infty} \sum_p \frac{f(p^l)^2}{p^l}.$$

are convergent.

Since this theorem is analogous to (I), there arises the question whether or not it is possible to replace (II) by a criterion (II') which relates to it as (I') does to (I). We shall prove that such is actually the case:

(II') *An additive $f(n)$ is almost periodic (B^2) if and only if its asymptotic distribution function $\sigma(x)$ possesses a second moment*

$$(11) \quad \int_{-\infty}^{+\infty} x^2 d\sigma(x) < \infty;$$

² P. Erdős and A. Wintner, "Additive functions and almost periodicity (B^2)," *American Journal of Mathematics*, vol. 62 (1940), pp. 635-645.

in which case one necessarily has

$$(12) \quad M\{f^2\} = \int_{-\infty}^{+\infty} x^2 d\sigma(x).$$

It is understood that $M\{g\}$ denotes the mean

$$(13) \quad M\{g\} = \lim_{n \rightarrow \infty} \frac{1}{n} (g(1) + g(2) + \cdots + g(n)),$$

if this limit exists. Incidentally, it will be clear from the proof of (II') that the condition (11) of almost periodicity (B^2) is satisfied if and only if f has an asymptotic distribution function and is such that

$$(14) \quad \bar{M}\{f^2\} < \infty,$$

where $\bar{M}\{g\}$ denotes the upper mean

$$(15) \quad \bar{M}\{g\} = \limsup_{n \rightarrow \infty} \frac{1}{n} (g(1) + g(2) + \cdots + g(n)), \text{ if } g \geq 0.$$

2 bis. It is very striking that the moment criterion (11) of (II') can insure almost periodicity (B^2). In fact, there will be given at the end of § 4 bis an example of a series of independent almost periodic (B^2) functions, with the property that the series is convergent everywhere to a limit function for which the square mean is infinite, although this function possesses an asymptotic distribution which is represented by the corresponding infinite convolution and which has a finite second moment. This example shows that the possibility of replacing (II) by (II') depends essentially on the properties of prime numbers, and not merely on the statistical independence of the terms of (1).

Similar remarks hold for the equivalence of (I) and (I').

3. For a given distribution function $\rho(x)$, $-\infty < x < +\infty$, and for a given positive integer i , let $\mu_i(\rho)$ denote the i -th moment

$$(16) \quad \mu_i(\rho) = \int_{-\infty}^{+\infty} x^i d\rho(x) \quad (\text{if } \int_{-\infty}^{+\infty} |x|^i d\rho(x) < \infty).$$

Thus, ρ has a finite standard deviation if and only if $\mu_2(\rho) < \infty$; in which case the square of the standard deviation of ρ is

$$(17) \quad v(\rho) = \mu_2(\rho) - \mu_1(\rho)^2 \geq 0.$$

Before proving (II'), it will be convenient to establish the following theorem:

(II*) An additive $f(n)$ is almost periodic (B^2) if and only if the asymptotic distribution functions (7) of its terms (3) are such as to make both numerical series

$$(18_1) \quad \sum_{k=1}^{\infty} \mu_1(\sigma_k) \qquad (18_2) \quad \sum_{k=1}^{\infty} \nu(\sigma_k)$$

convergent; in which case (18₁), (18₂) necessarily represent $\mu_1(\sigma)$, $\nu(\sigma)$, respectively, where σ is the asymptotic distribution function (9) of $f(n)$.

Notice that the convergence of (18₂) implies that $\nu(\sigma_k) < \infty$, i.e. $\mu_2(\sigma_k) < \infty$, for every k .

For a given distribution function $\rho(x)$, $-\infty < x < +\infty$, let ρ' , ρ'' denote the non-negative numbers

$$(19) \quad \rho' = \rho(-1), \quad \rho'' = 1 - \rho(1),$$

and put

$$(20) \quad \bar{\rho}(x) = \begin{cases} 0, & \text{if } -\infty < x \leq -1, \\ \rho(x) - \rho', & \text{if } -1 < x \leq 0, \\ \rho(x) + \rho'', & \text{if } 0 < x < 1, \\ 1, & \text{if } 1 \leq x < +\infty; \end{cases}$$

(so that $\bar{\rho}(x)$, $-\infty < x < +\infty$, obviously is a distribution function). Then one can express the theorem which relates to (I') in the same way as (II*) does to the equivalent formulation (II') of (II), as follows:

(I*) An additive $f(n)$ has an asymptotic distribution function if and only if the asymptotic distribution functions (7) of its terms (3) are such as to make the three numerical series

$$(21) \quad \sum_{k=1}^{\infty} (\sigma_k' + \sigma_k''), \quad \sum_{k=1}^{\infty} \mu_1(\bar{\sigma}_k), \quad \sum_{k=1}^{\infty} \nu(\bar{\sigma}_k)$$

convergent.

In fact, it is known³ that an infinite convolution $\sigma_1 * \sigma_2 * \dots$ is convergent if and only if the three numerical series (21) are convergent. Hence, (I*) follows from (I').

4. In order to avoid an interruption of the proofs, there will first be proved a relation between the existence of time averages and space averages of an arbitrary function $g(t)$, $0 \leq t < \infty$, which is almost periodic (B^λ) for some fixed $\lambda \geq 1$. The case of number-theoretical functions $f(n)$ may be

³ B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88, Theorem 34.

thought of as the particular case in which $g(t) = f(n)$ for $n - 1 < t \leq n$, where $n = 1, 2, \dots$ (and $f(0) = 0$, say). As will be seen in § 4 bis, the theorem to be obtained is not obvious in itself. In fact, it is to the effect that, in a case of almost periodicity, the general inequality of Fatou becomes an equality.

THEOREM. *If $\sigma(x)$, $-\infty < x < +\infty$, denotes the asymptotic distribution function of a real-valued function $g(t)$, $0 \leq t < \infty$, which is almost periodic (B^λ) for a fixed $\lambda \geq 1$, then*

$$(22) \quad M\{|g|^\mu\} = \int_{-\infty}^{+\infty} |x|^\mu d\sigma(x)$$

for every positive exponent $\mu \leq \lambda$ (which implies that

$$(23) \quad M\{g^\mu\} = \int_{-\infty}^{+\infty} x^\mu d\sigma(x)$$

for every positive $\mu \leq \lambda$, if μ is an integer).

Proof. For a fixed $T > 0$, let $\sigma_T(x)$, $-\infty < x < +\infty$, denote the distribution function of the function (of class L^λ) which is equal to $g(t)$ for $0 \leq t < T$ and has the period T . Then, by the definition of the asymptotic distribution function $\sigma(x)$ of $g(t)$,

$$(24) \quad \sigma_T \rightarrow \sigma \text{ as } T \rightarrow \infty$$

holds at every continuity point x of σ ; while obviously

$$\int_{-\infty}^{+\infty} |x|^\lambda d\sigma_T(x) = \frac{1}{T} \int_0^T |g(t)|^\lambda dt.$$

Since, by Fatou's inequality,

$$\int_{-\infty}^{+\infty} |x|^\lambda d \lim_{T \rightarrow \infty} \sigma_T(x) \leq \liminf_{T \rightarrow \infty} \int_{-\infty}^{+\infty} |x|^\lambda d\sigma_T(x),$$

it follows that

$$(25) \quad \int_{-\infty}^{+\infty} |x|^\lambda d\sigma(x) \leq M\{|g|^\lambda\} < \infty.$$

On the other hand, it is known⁴ that if $g_X(t)$, $0 \leq t < \infty$, denotes, for a fixed positive number X , the function which is equal to $-X$, $g(t)$ or X

⁴ A. S. Besicovitch, *Almost Periodic Functions*, Cambridge, 1932, p. 100.

according as $f(t) < -X$, $|g(t)| \leq X$, or $g(t) > X$, then g_X is almost periodic (B^λ) and one can find an $X_\epsilon > 0$ such that

$$(26) \quad M\{|g - g_X|^\lambda\} < \epsilon, \text{ if } X \geq X_\epsilon.$$

In view of (25), one can choose X_ϵ so large that

$$(27) \quad \int_{-\infty}^{-X} |x|^\lambda d\sigma(x) + \int_X^{+\infty} |x|^\lambda d\sigma(x) < \epsilon, \text{ if } X \geq X_\epsilon.$$

Since σ is monotone, one can assume that $x = \pm X_\epsilon$ are continuity points of $\sigma(x)$. Then it is clear from (24) that if $\epsilon > 0$ is given and X denotes X_ϵ , one can choose a positive $T = T(X, \epsilon) \equiv T_\epsilon$ so large that

$$(28) \quad |\sigma_T(x) - \sigma(x)| < \epsilon/X^\lambda \text{ for } x = \pm X, \quad (X = X_\epsilon).$$

In addition, $T = T_\epsilon$ may be so chosen so large that

$$(29) \quad \left| \int_{-X}^X |x|^\lambda d\sigma_T(x) - \int_{-X}^X |x|^\lambda d\sigma(x) \right| < \epsilon, \quad (X = X_\epsilon);$$

(this is clear from (24) and from Helly's term-by-term integration theorem, since $X = X_\epsilon$ is fixed). Furthermore, (26) assures that, since $X = X_\epsilon$ is fixed, one can choose $T = T_\epsilon$ so large that

$$(30) \quad \frac{1}{T} \int_0^T |g(t) - g_X(t)|^\lambda dt < \epsilon, \quad (X = X_\epsilon).$$

Finally, since $M\{|g|^\lambda\}$ exists ($< \infty$), one has

$$(31) \quad \left| \frac{1}{T} \int_0^T |g(t)|^\lambda dt - M\{|g|^\lambda\} \right| < \epsilon,$$

if $T = T_\epsilon$ is chosen sufficiently large.

Since the definitions of $\sigma_T(x)$ and $g_X(t)$ obviously imply that

$$\int_{-X}^X |x|^\lambda d\sigma_T(x) = \frac{1}{T} \int_0^T |g_X(t)|^\lambda dt - X^\lambda [\sigma_T(-X) + 1 - \sigma_T(X)],$$

it is clear from (27), (29) and (28) that

$$\left| \int_{-\infty}^{+\infty} |x|^\lambda d\sigma(x) - \frac{1}{T} \int_0^T |g_X(t)|^\lambda dt \right| < 4\epsilon + X^\lambda [\sigma(-X) + 1 - \sigma(X)];$$

so that, since $X^\lambda[\sigma(-X) + 1 - \sigma(X)]$ is certainly not larger than the sum of the two integrals on the left of (27), one has

$$(32) \quad \left| \int_{-\infty}^{+\infty} |x|^\lambda d\sigma(x) - \frac{1}{T} \int_0^T |g(t)|^\lambda dt \right| < 4\epsilon + \epsilon + \eta,$$

where

$$\eta = \left| \frac{1}{T} \int_0^T |g(t)|^\lambda dt - \frac{1}{T} \int_0^T |g_X(t)|^\lambda dt \right|.$$

Since it is seen from (30) and from Hölder's inequality that this η is majorized by a function of ϵ which tends to 0 as $\epsilon \rightarrow 0$, it follows from (31) and (32) that (22) is true for $\mu = \lambda$.

It is clear from this proof of (22) for $\mu = \lambda$, that (22) holds a fortiori for $0 < \mu < \lambda$, and that (23) is valid for every positive integer μ not greater than λ ; so that the proof of the Theorem is complete.

4 bis. That in the Theorem, just proved, the assumption of almost periodicity cannot be replaced by a mere average assumption, is shown by the following example:

In terms of a sequence of non-negative numbers a_1, a_2, \dots , define a function $g(t)$, $0 \leq t < \infty$, by placing $g(t) = 0$ for every t not contained in any of the intervals $n \leq t < n + n^{-1}$, where $n = 1, 2, \dots$, and $g(t) = a_n$ if t is in the n -th of these intervals. Clearly, the asymptotic distribution function, $\sigma(x)$, of $g(t)$ exists for every choice of the sequence $\{a_n\}$; in fact, $\sigma(x) = \frac{1}{2}(1 + \operatorname{sgn} x)$, so that the Stieltjes integral on the right of (16) exists and vanishes for every value of the exponent μ . On the other hand, one can choose the sequence $\{a_n\}$ so that (i) there exists a finite non-vanishing mean value $M\{g^2\}$; (ii) the mean value $M\{g^2\} = +\infty$; (iii) there does not exist a mean value $M\{g^2\} \leq +\infty$. In order to see this, it is sufficient to choose

$$(i) \quad a_n = n^{\frac{1}{2}}; \quad (ii) \quad a_n = n; \quad (iii) \quad a_n = \begin{cases} n, & \text{if } n = 2^m; \\ 0, & \text{if } n \neq 2^m. \end{cases}$$

In all three cases, (22) fails to hold for $\mu = 2$, although the integral on the right exists for $\mu = 2$.

In order to obtain a series of the type mentioned in § 2 bis, it is sufficient to consider the function $g(t)$, $0 \leq t < \infty$, defined by the convergent series $g_1(t) + g_2(t) + \dots$, where $g_n(t)$ has, for a fixed n , the value a_n or 0 according as t is or is not in the interval $n \leq t < n + n^{-1}$.

5. Besides the theorem proved in § 4, a parallel theorem on infinite convolutions will be needed.

LEMMA. *If an infinite convolution $\sigma_1 * \sigma_2 * \dots$ converges to a distribution function σ for which $\mu_2(\sigma) < \infty$, then the series (18₁), (18₂) are convergent and represent $\mu_1(\sigma)$, $\nu(\sigma)$, respectively.*

The proof of this Lemma will depend on the following known criterion:⁵

If a sequence of distribution functions $\sigma_1, \sigma_2, \dots$ is such as to make the series (18₂) convergent (so that, in particular, $\mu_2(\sigma_k) < \infty$ for $k = 1, 2, \dots$), then the infinite convolution $\sigma_1 * \sigma_2 * \dots$ is convergent if and only if the series (18₁) is convergent; in which case the distribution function $\sigma = \sigma_1 * \sigma_2 * \dots$ has a second moment $\mu_2(\sigma) < \infty$, and the series (18₁), (18₂) represent $\mu_1(\sigma)$, $\nu(\sigma)$, respectively.

It is easily verified from the definition of the convolution $\alpha * \beta$ of two distribution functions $\alpha = \alpha(x)$, $\beta = \beta(x)$, that

$$(33) \quad \mu_2(\alpha * \beta) < \infty \text{ if and only if } \mu_2(\alpha) + \mu_2(\beta) < \infty.$$

Furthermore,

$$(34) \quad \text{if } \mu_2(\alpha * \beta) < \infty, \text{ then } \nu(\alpha * \beta) = \nu(\alpha) + \nu(\beta).$$

The Lemma may now be proved as follows:

Suppose that $\mu_2(\nu) < \infty$ holds for a given convergent infinite convolution $\sigma = \sigma_1 * \sigma_2 * \dots$. Then repeated application of (33) and (34) shows that

$$(35) \quad \nu(\sigma_1) + \dots + \nu(\sigma_k) + \nu(\sigma_{k+1} * \sigma_{k+2} * \dots) = \nu(\sigma)$$

holds for every k . In fact, it is known⁶ that if $\sigma = \sigma_1 * \sigma_2 * \dots$ is a convergent infinite convolution, then the infinite convolution $\sigma_{k+1} * \sigma_{k+2} * \dots$, where k is arbitrarily fixed, converges to a distribution function, and that the convolution of this distribution function and of $\sigma_1 * \sigma_2 * \dots * \sigma_k$ is σ .

Since (35), (17) and the assumption $\mu_2(\sigma) < \infty$ imply the convergence of the series (18₂), the criterion quoted immediately after the Lemma shows that the proof of the Lemma is complete.

6. Next, it will be shown that, in virtue of the representation (7) of

⁵ This criterion is implied by the proof, though not the wording, of Theorems 4 and 5 in the paper of B. Jessen and A. Wintner, *loc. cit.*³, pp. 56-58. The method applied there is that of the Fourier-Stieltjes transforms. For the above wording and for a proof which does not make use of Fourier-Stieltjes transforms, cf. E. R. van Kampen, "Infinite product measures and infinite convolutions," *American Journal of Mathematics*, vol. 62 (1940), pp. 417-448; more particularly, (9) on p. 442.

⁶ B. Jessen and A. Wintner, *loc. cit.*³, Theorem 2.

the asymptotic distribution functions $\sigma_1(x)$, $\sigma_2(x)$, \dots of the terms $f_1(n)$, $f_2(n)$, \dots of an arbitrary additive $f(n)$, the two series (10_1) , (10_2) are convergent if and only if the two series (18_1) , (18_2) are convergent.

It is clear from (7) and (16) that if k is fixed and $p = p_k$ denotes the k -th prime number, then

$$(36_1) \quad \mu_1(\sigma_k) = \frac{p}{p-1} \sum_{l=1}^{\infty} \frac{f(p^l)}{p^l}; \quad (36_2) \quad \mu_2(\sigma_k) = \frac{p}{p-1} \sum_{l=1}^{\infty} \frac{f(p^l)^2}{p^l},$$

provided that the series (36_1) , (36_2) are absolutely convergent, where it is understood that if the non-negative series (36_2) is divergent, then $\mu_2(\sigma_k) = \infty$. Furthermore, it is clear from the Schwarz inequality,

$$(37) \quad \left(\sum_{l=1}^{\infty} \frac{f(p^l)}{p^l} \right)^2 \leq \frac{1}{p-1} \sum_{l=1}^{\infty} \frac{f(p^l)^2}{(p^{3l})^2}, \text{ where } \frac{1}{p-1} = \sum_{l=1}^{\infty} \frac{1}{(p^{3l})^2},$$

that the absolute convergence of the series (36_1) is implied by the convergence of the series (36_2) . Consequently, both relations (36_1) , (36_2) hold for a fixed k whenever $\mu_2(\sigma_k) < \infty$. And (16), (17) show that $\mu_2(\sigma_k) < \infty$ is equivalent to $\nu(\sigma_k) < \infty$ (a condition which is certainly satisfied for every k whenever (18_2) is a convergent series). Suppose that $\nu(\sigma_k) < \infty$ for every k .

It is clear from (36_1) , (36_2) and (37) that

$$\mu_1(\sigma_k)^2 \leq \frac{p}{(p-1)^2} \mu_2(\sigma_k); \text{ so that } \nu(\sigma_k) \geq \left(1 - \frac{p}{(p-1)^2}\right) \mu_2(\sigma_k),$$

by the definition (17) of ν . On substituting $\mu_2(\sigma_k)$ from (36_2) into the last inequality, one obtains

$$(38) \quad \nu(\sigma_k) \geq C_p \sum_{l=1}^{\infty} \frac{f(p^l)^2}{p^l}, \text{ where } C_p = \frac{p^3 - 3p^2 + p}{(p-1)^3} \rightarrow 1 \text{ as } p = p_k \rightarrow \infty.$$

On the other hand, since (17) implies that $\nu(\sigma_k) \leq \mu_2(\sigma_k)$, one sees from (36_2) that

$$(39) \quad \nu(\sigma_k) \leq A_p \sum_{l=1}^{\infty} \frac{f(p^l)^2}{p^l}, \text{ where } A_p = \frac{p}{p-1} \rightarrow 1 \text{ as } p = p_k \rightarrow \infty.$$

The relations (38) and (39) imply that either both series (10_2) , (18_2) are convergent or both are divergent. It follows that in order to complete the proof of the last italicized statement, it is sufficient to prove that if the series (10_2) , (18_2) are convergent, then either both series (10_1) , (18_1) are convergent or both are divergent.

To this end, notice first that, since $ab \leq a^2 + b^2$,

$$(40) \quad \sum_p \sum_{l=2}^{\infty} \frac{|f(p^l)|}{p} \leq \sum_p \sum_{l=2}^{\infty} \frac{f(p^l)^2}{p^l} + \sum_p \sum_{l=2}^{\infty} \frac{1}{p^l}.$$

But the first double series on the right of (40) is majorized by the series (10₂), which is supposed to be convergent; so that, since

$$\sum_p \sum_{l=2}^{\infty} \frac{1}{p^l} = \sum_p \frac{1}{p(p-1)} < \infty,$$

the double series on the left of (40) is convergent. It follows, therefore, from (36₁) that the series (18₁) is convergent if and only if the series

$$(41) \quad \sum_p \frac{p}{p-1} \frac{f(p)}{p} = \sum_p \frac{f(p)}{p-1}, \quad (l=1),$$

is convergent.

Hence, the statement that either both series (18₁), (10₁) are convergent or both are divergent is equivalent to the statement that either both series (41), (10₁) are convergent or both are divergent. Since the difference of the series (41), (10₁) is

$$(42) \quad \sum_p \frac{f(p)}{p(p-1)},$$

it follows that all that remains to be shown is that the series (42) is convergent if either (41) or (10₁) is a convergent series. But both $\{p^{-1}\}$ and $\{(p-1)^{-1}\}$ are monotone and bounded sequences of numbers (which tend, in fact, to 0). Hence, it is seen from a standard convergence criterion (partial summation), that the convergence of either of the series (41), (10₁) implies the convergence of the series (42).

This completes the proof of the last italicized statement.

7. The proofs of (II*), § 3 and (II'), § 2 are now immediate. In fact, it is clear from the Lemma of § 5 and the Theorem of § 4 (where $\lambda = 2 = \mu$), that, because of (I'), § 1 and the italicized result of § 6, both (II*), § 3 and (II'), § 2 are equivalent to (II), § 2.

QUEENS COLLEGE,
THE JOHNS HOPKINS UNIVERSITY.

ON THE ALMOST PERIODICITY OF ADDITIVE NUMBER-THEORETICAL FUNCTIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

1. By an additive function $f = f(n)$ is meant a sequence $f(1), f(2), \dots$ defined for every positive integer n in such a way that

$$(1) \quad f(n_1 n_2) = f(n_1) + f(n_2) \text{ whenever } (n_1, n_2) = 1; \quad (f(1) = 0).$$

Thus, if $p = p_k$ denotes the k -th prime number and l is a positive integer, the correspondence $a_{lk} = f(p_k^l)$ establishes a one-to-one correspondence between arbitrary additive functions f and arbitrary double sequences of numbers $\{a_{lk}\}$. With every additive function $f(n)$, there is associated the sequence $f_1(n), f_2(n), \dots$ of additive functions, where the double sequence $\{f_j(p_k^l)\}$ of $f_j(n)$ is defined as follows:

$$(2) \quad f_j(p_k) = \begin{cases} f(p_k^l) & \text{if } k = 1, 2, \dots, j, \\ 0 & \text{if } k > j, \end{cases} \quad (l = 1, 2, \dots).$$

It is known¹ that the real additive function $f(n)$ has an asymptotic distribution function if and only if both series

$$(3_1) \quad \sum_p \frac{f^+(p)}{p}; \quad (3_2) \quad \sum_p \frac{f^+(p)^2}{p}$$

are convergent, where $y^+ = f^+$ is defined by placing

$$(4) \quad y^+ = y \text{ or } y^+ = 1 \text{ according as } |y| < 1 \text{ or } |y| \geq 1.$$

It is also known² that an additive function $f(n)$ is almost periodic (B) if and only if both series

$$(5_1) \quad \sum_p \frac{f(p)}{p}; \quad (5_2) \quad \sum_{l=1}^{\infty} \sum_p \frac{|f(p^l)|^2}{p^l}$$

are convergent.

By a suitable modification of the proof of the latter theorem, it will be shown in the present paper that an additive function $f(n)$ is almost periodic (B) if and only if the four series

* Received April 8, 1940.

¹ P. Erdős and A. Wintner, "Additive arithmetical functions and statistical independence," *American Journal of Mathematics*, vol. 61 (1939), pp. 713-721.

² P. Erdős and A. Wintner, "Additive functions and almost periodicity (B²)," *American Journal of Mathematics*, vol. 62 (1940), pp. 635-645.

$$(6_1) \sum_p \frac{f(p)}{p}; \quad (6_2) \sum_p \frac{|f^*(p)|^2}{p}; \quad (6_3) \sum_{l=2}^{\infty} \sum_p \frac{|f(p^l)|}{p^l}; \quad (6_4) \sum_{|f(p)| \geq 1} \frac{|f(p)|}{p}$$

are convergent.

Notice that the exterior summation index runs in (5₂) and (6₃) from $l=1$ and $l=2$, respectively.

If $f = f_I + if_{II}$, where f_I, f_{II} are real, then f is additive and almost periodic (B) if and only if so are both functions f_I, f_{II} . Furthermore, since (4) implies that

$$|f^*|^2 \leq (f_I^*)^2 + (f_{II}^*)^2 \leq 2|f^*|^2, \quad (f = f_I + if_{II}),$$

the 4 series (6₁), (6₂), (6₃), (6₄) are convergent for $f = f_I + if_{II}$ if and only if so are the 4 + 4 series which one obtains by writing f_I and f_{II} for f . Hence, it is sufficient to prove the italicized theorem for the case of a real-valued f . This restriction will always be assumed.

2. In order to prove first the sufficiency of the criterion of the italicized theorem, suppose that the four series (6₁)–(6₄) are convergent.

Let $F(n)$ be the additive function for which the double sequence $\{\{F(p_k^l)\}\}$ is given by

$$(7) \quad F(p^l) = \begin{cases} f(p^l) & \text{if } |f(p)| \geq 1, \\ f(p^l) - f(p) & \text{if } |f(p)| < 1, \end{cases} \quad (p = p_k).$$

Since the proof given *loc. cit.*² (beginning of § 5) for

$$(8) \quad \sum_{l=1}^{\infty} \sum_p \frac{|F(p^l)|}{p^l} < \infty$$

on the assumption of the convergence of the series (5₂) actually uses the convergence (6₃) and (6₄) only, (8) is satisfied. Hence,

$$(9) \quad \lim_{j \rightarrow \infty} \sum_{l=1}^{\infty} \sum_{\substack{p|m \\ p > j}} \frac{|F(p^l)|}{p^l} = 0.$$

Since obviously

$$\sum_{m=1}^n \sum_{\substack{p|m \\ p > j}} |F(p^l)| \leq \sum_{l=1}^{\infty} \sum_{p > j} \frac{n}{p^l} |F(p^l)|$$

for every j , it follows from (9) that

$$(10) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n \sum_{\substack{p|m \\ p > j}} |F(p^l)| \rightarrow 0 \text{ as } j \rightarrow \infty.$$

But if $F_j(n)$ denotes the function which belongs to $F(n)$ in the same way

as the function $f_j(n)$, which is defined by (2), belongs to $f(n)$, then, since F and F_j are additive functions of n ,

$$\frac{1}{n} \sum_{m=1}^n |F(m) - F_j(m)| \leq \frac{1}{n} \sum_{m=1}^n \sum_{\substack{p|m \\ p > j}} |F(p^t)|.$$

Hence, (10) implies that

$$(11) \quad \bar{M}\{|F - F_j|\} \rightarrow 0 \text{ as } j \rightarrow \infty,$$

where $\bar{M}\{g\}$ denotes the upper mean value

$$\bar{M}\{g\} = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n g(m)$$

of a non-negative function g of n .

Let $G(n)$ denote the additive function

$$(12) \quad G = f - F.$$

Then (7) implies that, for every prime p ,

$$(13) \quad G(p) = \begin{cases} 0 & \text{if } |f(p)| \geq 1; \\ f(p) & \text{if } |f(p)| < 1; \end{cases}$$

so that

$$\sum_p \frac{G(p)^2}{p} = \sum_{|f(p)| < 1} \frac{f(p)^2}{p}.$$

Since the series on the right is, in view of (4), majorized by the series (6₂), which is supposed to be convergent, it follows that

$$(14) \quad \sum_p \frac{G(p)^2}{p} < \infty.$$

On the other hand, it is clear from (13) and from the convergence of the series (6₁) and (6₄), that

$$(15) \quad \sum_p \frac{G(p)}{p} \text{ is convergent,}$$

since

$$\sum_p \frac{f(p)}{p} = \sum_p \frac{G(p)}{p} + \sum_{|f(p)| \geq 1} \frac{f(p)}{p}.$$

But (14) and (15) imply, as shown *loc. cit.*², § 8-§ 8 bis, that

$$(16) \quad \bar{M}\{|G - G_j|^2\} \rightarrow 0 \text{ as } j \rightarrow \infty,$$

where $G_j(n)$ denotes the additive function which belongs to $G(n)$ in the same way as the function $f_j(n)$, which is defined by (2), belongs to $f(n)$.

Since (12) obviously implies that $G_j = f_j - F_j$, it is clear that

$$\bar{M}\{|f - f_j|\} \leq \bar{M}\{|F - F_j|\} + \bar{M}\{|G - G_j|\};$$

it follows therefore from (11), (16) and from an application of the Schwarz inequality to $\bar{M}\{|G - G_j|\}$, that

$$(17) \quad \bar{M}\{|f - f_j|\} \rightarrow 0 \text{ as } j \rightarrow \infty.$$

Finally, it is known³ that if an additive function $g(n)$ is, with reference to a fixed prime number $p = p_j$, such that

$$(18) \quad g(n) = 0 \text{ whenever } p_j \nmid n,$$

then $g(n)$ is almost periodic (B) if and only if

$$(19) \quad \sum_{l=1}^{\infty} \frac{|g(p^l)|}{p} < \infty.$$

But it is clear from the definition (2) of the additive functions f_j of n , that the additive function $f^{(j)}$ of n which is defined for $j = 1, 2, \dots$ by

$$(20) \quad f^{(1)}(n) = f_1(n), \quad f^{(2)}(n) = f_2(n) - f_1(n), \quad f^{(3)}(n) = f_3(n) - f_2(n), \dots$$

is such that (18) is satisfied by $g = f^{(j)}$. Furthermore, it is seen from the definitions (2), (20) of the additive function $f^{(j)}$ of n , that the convergence of the series (6₃) implies that

$$\sum_{l=1}^{\infty} \frac{|f^{(j)}(p^l)|}{p^l} < \infty \text{ for every } p = p_j.$$

This means that (19) is satisfied by $g = f^{(j)}$ for every j . Consequently, $f^{(j)}$ is almost periodic (B). Since (20) implies that

$$f_j = f^{(1)} + f^{(2)} + \dots + f^{(j)},$$

and since the functions which are almost periodic (B) form a linear space, it follows that the additive function f_j of n is almost periodic (B) for every j . Hence, it is clear from (17) that the proof for the almost periodicity (B) of $f(n)$ is now complete.

3. This proves that the convergence of the four series is a sufficient condition for the almost periodicity (B) of f . In order to prove the necessity of this condition, suppose that $f(n)$ is a given real additive function which is almost periodic (B).

Since $f(n)$ then has an asymptotic distribution function, both series

³ E. R. van Kampen and A. Wintner, "On the almost periodic behavior of multiplicative number-theoretical functions," *American Journal of Mathematics*, vol. 62 (1940), pp. 613-626, Theorem II, $\lambda = 1$.

(3₁), (3₂) are convergent. And, in view of (4), the convergence of (3₂) implies that

$$(21) \quad \sum_{|f(p)| \geq 1} \frac{1}{p} < \infty.$$

In terms of the given $f(n)$, define an additive function $D(n)$ by placing

$$(22) \quad D = f - H,$$

where $H = H(n)$ denotes that additive function for which the double sequence $\{H(p_k^l)\}$ is given by

$$(23) \quad H(p^l) = \begin{cases} f(p^l), & \text{if } l \neq 1, \\ f(p) & \text{if } l = 1 \text{ and } |f(p)| \geq 1, \\ 0 & \text{if } l = 1 \text{ and } |f(p)| < 1. \end{cases}$$

Accordingly,

$$D(p^l) = \begin{cases} 0, & \text{if } l \neq 1, \\ 0, & \text{if } l = 1 \text{ and } |f(p)| \geq 1 \\ f(p), & \text{if } l = 1 \text{ and } |f(p)| < 1, \end{cases}$$

and so it is clear from (21) and from the convergence of the series (3₁) and (3₂), that one obtains four convergent series by writing D for f in (6₁)–(6₄). It follows, therefore, from the result proved in § 2, that $D(n)$ is almost periodic (B). Since f is almost periodic (B) by assumption, one sees from (22) that H is almost periodic (B). In particular,

$$(24) \quad \bar{M}\{|H|\} < \infty.$$

But (21) and (24) imply, after an obvious adaptation of the estimates carried out *loc. cit.*², § 11, that

$$(25) \quad \sum_{l=1}^{\infty} \sum_p \frac{|H(p^l)|}{p^l} < \infty,$$

this series (25) being the analogue of the last series *loc. cit.*², § 11, in case the almost periodic class (B^2) is replaced by (B).

It is now easy to prove the convergence of the four series (6₁)–(6₄). In fact, it is clear from (23) that (25) may be written in the form

$$(26) \quad \sum_{|f(p)| \geq 1} \frac{|f(p)|}{p} + \sum_{l=2}^{\infty} \sum_p \frac{|f(p^l)|}{p^l} < \infty.$$

Furthermore, both series (3₁), (3₂) are convergent in view of the existence of the asymptotic distribution function of the function f . Since (3₂) is identical with (6₂), while (26) implies the convergence of (6₃) and (6₄), the convergence of the three series (6₁), (6₂), (6₃) follows. Finally, since (21) and (26) imply the absolute convergence of the series

$$(27_1) \quad \sum_{|f(p)| \geq 1} \frac{1}{p}; \quad (27_2) \quad \sum_{|f(p)| \geq 1} \frac{f(p)}{p},$$

respectively, the convergence of (6_1) follows from the convergence of (3_1) and from the fact that, in view of (4), the series (6_1) may formally be written as the sum of the three series (3_1) , (27_1) , (27_2) .

4. A careful perusal of the proofs, applied *loc. cit.*² and above, shows that, by standard applications of the inequalities of Hölder and Minkowski, one can generalize the criteria (5_1) – (5_2) and (6_1) – (6_4) of the respective almost periodic classes (B^2) and $(B) = (B^1)$ as follows: *An additive function $f(n)$ is almost periodic (B^λ) for a fixed $\lambda \geq 1$ if and only if the four series*

$$(28_1) \sum_p \frac{f(p)}{p}; \quad (28_2) \sum_p \frac{|f^+(p)|^2}{p}; \quad (28_3) \sum_{l=2}^{\infty} \sum_p \frac{|f(p^l)|^\lambda}{p^l}; \quad (28_4) \sum_{|f(p)| \geq 1} \frac{|f(p)|^\lambda}{p}$$

are convergent. (Correspondingly, the convergence of (28_4) , (28_2) , (28_3) is equivalent to the convergence of the single series (5_2) , if $\lambda = 2$.)

A consequence of the italicized theorem is that if the double sequence $\{f(p_k^l)\}$ of an additive function $f(n)$ is bounded, then $f(n)$ either is almost periodic (B^λ) for arbitrarily large λ or is not even almost periodic $(B) = (B^1)$. This is not obvious in itself, since $f(n)$ is not in general a bounded function when its double sequence is bounded.

QUEENS COLLEGE,
THE JOHNS HOPKINS UNIVERSITY.

ON THE SPHERICAL APPROACH TO THE NORMAL DISTRIBUTION LAW.*

By PHILIP HARTMAN and AUREL WINTNER.

Introduction. There are two classical "geometrical" approaches to the normal distribution law. One of these is represented by the theory of the addition of independent random variables or, equivalently, by the theory of convolutions. This approach, followed in a general and precise manner by P. Lévy and his followers, is based on the consideration of a product distribution on an n -dimensional cube, a distribution which is then projected orthogonally on the principal diagonal of this cube.¹ The other approach, which is due to Boltzmann and is reproduced in some of Borel's elementary text-books on the calculus of probability, has as its starting point, not the theory of independent random variables, but rather the simplest model of the Maxwell theory of velocity distributions.² This approach, which plays a fundamental rôle in the investigations of P. Lévy³ and N. Wiener⁴ in functional analysis, is based on the consideration of the equidistribution on the surface of an n -dimensional sphere, a distribution which is then projected orthogonally on a diameter of this sphere.

If the unit of length is increased in the proportion $1:\sqrt{n}$ in case of the first approach, and decreased in the same proportion in the second approach, there results, as $n \rightarrow \infty$, a symmetric normal distribution in both cases.

It is known⁵ that the Fourier-Stieltjes transform of the equidistribution on the surface of an n -dimensional sphere of radius r is the Bessel function $J_{\frac{1}{2}n-1}^*(r|u|)/J_{\frac{1}{2}n-1}^*(0)$, where $J_\nu^*(z) = z^{-\nu}J_\nu(z)$, and that this Bessel function is also the Fourier-Stieltjes transform of the 1-dimensional distribution which represents the projection on a diameter. In fact,⁶ any 1-dimensional

* Received April 15, 1940.

¹ Cf., e.g., Lévy [11]. As to the geometrical interpretation of the convolution process by means of orthogonal projections, cf. Sommerfeld [17], where the simplest case of the "Abrundungsfehler" is considered.

² Cf. Boltzmann [2], vol. 2, pp. 96-100 and, e.g., Borel [3], pp. 44-50; also Borel [4], pp. 90-93; and Borel and Deltheil [5], pp. 134-136.

³ Lévy [12]; also Lévy [13].

⁴ Cf., e.g., Wiener [18], pp. 135-143.

⁵ Cf., e.g., Wintner [20], p. 313, where references are given to the principle of Huyghens; Jessen and Wintner [9], p. 59; Wintner [21]. Some of these things were recently rediscovered by Schoenberg (e.g., Schoenberg [15], Lemma 4); cf. also Blumenthal [1].

⁶ Cf., e.g., Jessen and Wintner [9], p. 55; Wintner [21], p. 76.

distribution which represents the projection of an n -dimensional distribution function of radial symmetry (i. e., a distribution which is built up by means of an arbitrary Stieltjes weight factor from spherical equidistributions belonging to varying r) has the same Fourier-Stieltjes transform as the projected distribution.

In view of the multi-dimensional ⁷ analogue of Lévy's inversion formula of Fourier-Stieltjes transforms, this spherically stratified decomposition of an arbitrary distribution function of radial symmetry into spherical equidistributions is known to be equivalent to the Cauchy-Poisson formula for spherical waves in n -dimensions.⁸

The results of the present paper concern certain questions connected partly with the above topics and partly with a problem ⁹ suggested by the most primitive approach to Maxwell's law of velocity distribution. If $\delta(v_x, v_y, v_z)$ denotes the density of probability at the point $v = (v_x, v_y, v_z)$ of the velocity space, then Maxwell's assumptions imply that δ is a function of $|v| = (v_x^2 + v_y^2 + v_z^2)^{\frac{1}{2}}$ alone, that the probability densities of each of the velocity components v_x, v_y, v_z depend on the respective components alone, and that the latter densities are represented, up to adjusting factors of proportionality, by the *same function* δ as the probability density of the speed $|v|$. In fact, this condition of the preservation of the density function under projections is obviously satisfied if $\log \delta(|v|)$ is proportional to $|v|^2$. There rises, therefore, the question whether or not this property of preservation of the probability density under projection is in itself sufficient to assure that $\delta(|v|)$ defines the Maxwell distribution. The result of § 5 will imply that the Maxwell law may be deduced from this functional condition alone.

§ 3 deals with the class of distribution functions which may be represented as stratifications of a given sheaf of distribution functions. The results obtained in this section are illustrated in § 4 by their application to the special case of stable distribution functions. The simplest and least restricted case of this particular case is the one where the underlying stable distribution is normal. This limiting case will be separately studied in § 2 by an elementary approach.

As to this approach, which in § 3 will be extended to the general case, a few methodical remarks seem to be of interest. Recently, Schoenberg ¹⁰ has rediscovered the above-mentioned Cauchy-Poisson decomposition and, in particular, the fact ¹¹ that the Fourier-Stieltjes transform of the spherical equi-

⁷ Haviland [8], I.

⁸ Cf., e. g., Wintner [20], pp. 316-319; Jessen and Wintner [9], p. 55; Wintner [21], p. 76.

⁹ For more refined approaches, cf. e. g., Boltzmann [2], vol. 1, chap. 1.

¹⁰ Schoenberg [15], p. 816; cf. Blumenthal [1]

¹¹ Schoenberg [14], p. 791; cf. Blumenthal [1].

distribution of radius r is $J_{\frac{1}{2}n-1}^*(r|u|)/J_{\frac{1}{2}n-1}^*(0)$. Actually, Schoenberg's considerations¹² concern all functions which are Fourier-Stieltjes transforms of radially symmetric n -dimensional distributions for every n . Since Borel's approach to the normal distribution law also has escaped Schoenberg, he rediscovers¹² the spherical approach to the Gaussian law by applying to the Bessel functions, mentioned above, the continuity theorem of Fourier-Stieltjes transforms, instead of proceeding directly as Boltzmann, or Borel and Lévy do.¹³ And he applies the same method to the spherical stratification formula of Cauchy-Poisson. Now, it will be seen in § 2 that the intuitive and elementary method may be transferred without much effort to this general case of an arbitrary Stieltjes factor of spherical stratification. In particular, the method of Fourier-Stieltjes transforms, which is so fundamental in most problems of mathematical statistics, turns out to be a ballast in the present case.

1. Let $\phi_n(E_n)$ denote a distribution function on the n -dimensional Euclidean space R_n , that is, ϕ_n is a completely additive, non-negative set function defined for all Borel sets E_n of R_n in such a way that $\phi(R_n) = 1$. It will be supposed that ϕ_n is radially symmetric, i. e., $\phi(E'_n) = \phi(E_n)$ if E'_n is the image of any Borel set E_n under an arbitrary rotation of the space R_n about the origin. For $k = 1, 2, \dots, n-1$, a k -dimensional radially symmetric distribution function $\phi_k(E_k)$ can be associated with $\phi_n(E_n)$ in the following manner:

Let R_k be a k -dimensional hyperplane through the origin of R_n , and E_k a Borel set on R_k , finally $P_n(E_k)$ the set of those points in R_n whose orthogonal projection on R_k is in E_k . Then a distribution function ϕ_k is defined by the relation¹⁴

$$(1) \quad \phi_k(E_k) = \phi_n(P_n(E_k)).$$

ϕ_k is called the k -dimensional projection of ϕ_n ; in virtue of the radial symmetry of ϕ_n , it is independent of the choice of the hyperplane R_k . It is clear from this definition that ϕ_k is also the k -dimensional projection of ϕ_j , for $j = k+1, \dots, n$.

The set functions ϕ_1, \dots, ϕ_n of radial symmetry may be replaced by the non-decreasing point functions $\rho_1(r), \dots, \rho_n(r)$ which are defined for $0 \leq r < \infty$ as follows:

$$(2_k) \quad \rho_k(r) = \phi_k(E_k^r), \text{ if } r > 0; \quad \rho_k(0) = 0,$$

¹² Schoenberg [15], pp. 816-821.

¹³ Incidentally, Schoenberg's [15] central formula (2.4), which is due to Laplace, may be found on p. 421 of Watson's Treatise on Bessel Functions.

¹⁴ Cf. Jessen and Wintner [9], p. 55; Wintner [21], p. 76.

where E_k^r denotes the k -dimensional sphere of radius r about the origin of R_k . In the case $k = 1$, the function ρ_1 is usually replaced by the symmetric distribution $\sigma(x)$, $-\infty < x < +\infty$,

$$(2 \text{ bis}) \quad \sigma(x) = \phi_1(E_1(x)), \quad -\infty < x < +\infty,$$

where $E_1(x)$ denotes the half-line $(-\infty, x)$ on R_1 . Obviously, these distribution functions ρ_k, σ satisfy the boundary conditions

$$(3_k) \quad \rho_k(0) = 0, \quad \rho_k(+\infty) = 1, \quad (\rho_k(r) = 0 \text{ for } -\infty < r < 0),$$

$$(3 \text{ bis}) \quad \sigma(-\infty) = 0, \quad \sigma(+\infty) = 1;$$

also, in virtue of the symmetry of ϕ_1 ,

$$(4) \quad \sigma(-x) = 1 - \sigma(x).$$

In the sequel, the functions ρ_k (where $k < n$) and σ also will be referred to as projections of ϕ_n .

It is clear that the relation between ρ_n and ρ_k is given by the formula

$$(5_{nk}) \quad \rho_k(r) = \rho_n(r) + B_n^k \int_r^\infty \left[\int_{\text{arc } \cos r/t}^{\frac{1}{2}\pi} (\cos \theta)^{k-1} (\sin \theta)^{n-k-1} d\theta \right] d\rho_n(t), \quad r > 0,$$

where

$$(6) \quad B_n^k = A_n \cdot A_{n-1} \cdots A_{k+1} / A_{n-k} \cdot A_{n-k-1} \cdots A_2,$$

and ¹⁵

$$(7) \quad A_j = \left[\int_0^\pi \sin^{j-2} \alpha d\alpha \right]^{-1} \sim (2\pi)^{-\frac{1}{2}j\frac{1}{2}},$$

since the integral $B_n^k \int (\cos \theta)^{k-1} (\sin \theta)^{n-k-1} d\theta$ in (5_{nk}) is that portion of the $(n-1)$ -dimensional area of the boundary of the sphere of radius t ($> r$), whose projection on the hyperplane R_k is on the sphere E_k^r . The relation between σ and ρ_n is given by

$$(8) \quad \sigma(x) = \frac{1}{2} + \frac{1}{2}\rho_1(x), \quad \text{where } x > 0; \text{ cf. (4).}$$

The formula (5_{nk}) may be rewritten by introducing the distribution function

$$(9) \quad \psi_{nk}(r) = B_n^k \int_{\text{arc } \cos r}^{\pi/2} (\cos \theta)^{k-1} (\sin \theta)^{n-k-1} d\theta, \quad \text{if } 0 \leq r \leq 1;$$

$$\psi_{nk}(r) = 1, \quad \text{if } r > 1,$$

(a function which obviously bears the same relationship to the k -dimensional projection of the n -dimensional spherical equidistribution of radius 1, as the function (2_k) does to ϕ_k). The formula (5_{nk}) then becomes

¹⁵ Cf., e. g., Borel and Deltheil [5], p. 135 and p. 187.

$$(10) \quad \rho_k(r) = \int_0^\infty \psi_{nk}(r/t) d\rho_n(t), \quad r > 0.$$

In view of (2_k), the formula (10) represents, in terms of the arbitrary Stieltjes weight factor $\rho_n(t)$, the stratified decomposition of an arbitrary n -dimensional distribution function of radial symmetry into equidistributions on surfaces of spheres of varying radii.

While it is obvious that the distribution functions ρ_k and ϕ_k determine each other uniquely, and that ρ_k is uniquely determined by ρ_n , it is not so obvious that ρ_n is uniquely determined by ρ_k . That such is the case, nevertheless, may be seen by considering formula (10) as a convolution on a logarithmic scale; the uniqueness of ρ_n then follows from the uniqueness theorem of Fourier-Stieltjes transforms. This remark, depending on Fourier-Stieltjes transforms, is not used in the sequel.

For the sake of brevity, a 1-dimensional distribution function which is the projection of an n -dimensional radially symmetric distribution function for arbitrarily large n , will be called a distribution function of class Ω .

2. On the basis of the elementary geometrical relations collected above, it is easy to prove that a *distribution function* $\sigma(x)$, $-\infty < x < +\infty$, is of class Ω if and only if there exists a distribution function $\tau(t)$, $-\infty < t < +\infty$, such that

$$(11) \quad \tau(0) = 0, \quad \tau(+\infty) = 1$$

and

$$(12) \quad \sigma(x) = \int_0^\infty \sigma^*(x/t) d\tau(t),$$

where $\sigma^*(x)$ is the symmetric normal distribution function, of unit standard deviation, i. e.,

$$(13) \quad \sigma^*(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy.$$

In order to prove this, suppose first that $\sigma(x)$ is a distribution function of class Ω . Then there exists a $\phi_n(E_n)$ and corresponding functions (2_n), (2₁), such that (5_{n1}) and (8) hold for $n = 1, 2, \dots$. If n is fixed, these relations may be rewritten in the form

$$(14) \quad \sigma(x) = \frac{1}{2} + \frac{1}{2}\rho_n(x) + A_n n^{-\frac{1}{2}} \int_{n^{-\frac{1}{2}}x}^\infty \int_0^{x/t} (1 - y^2/n)^{\frac{1}{2}(n-3)} dy d\rho_n(n^{\frac{1}{2}}t), \quad x > 0,$$

if one changes the integration variables from θ to $y = n^{\frac{1}{2}} \cos \theta$ and from t to $n^{-\frac{1}{2}}t$.

According to the selection theorems of Helly,¹⁶ there exists a non-decreasing function $\tau(t)$ and an increasing sequence $\{m_n\}$ of positive integers, such that, as $n \rightarrow \infty$,

$$(15) \quad \rho_{m_n}(m_n^{\frac{1}{2}}t) \rightarrow \tau(t),$$

where the sign \rightarrow is meant in the sense of theory of monotone functions, (i. e., in the sense that one has convergence at every continuity point t of the limit function τ). It is clear from (15) that

$$(16) \quad \rho_{m_n}(x) \rightarrow \tau(+0), \text{ for all } x > 0, \text{ as } n \rightarrow \infty.$$

In view of the term-by-term integration theorem of Helly,¹⁶ it is also clear from (15) that

$$(17) \quad \int_{\epsilon}^{\infty} \sigma^*(x/t) d\rho_{m_n}(m_n^{\frac{1}{2}}t) \rightarrow \int_{\epsilon}^{\infty} \sigma^*(x/t) d\tau(t),$$

if $x > 0$ is fixed and ϵ is an arbitrary positive number such that $t = \epsilon$ is a continuity point of $\tau(t)$. On the other hand, since

$$(18) \quad A_n n^{-\frac{1}{2}} \rightarrow (2\pi)^{-\frac{1}{2}},$$

holds in virtue of (7), and since

$$(1 - y^2/n)^{\frac{1}{2}(n-3)} \rightarrow e^{-\frac{1}{2}y^2},$$

holds uniformly for $|y| \leq \text{const.}$, where const. is arbitrary but fixed, one sees, by choosing $\text{const.} = x/\epsilon$, that

$$A_n n^{-\frac{1}{2}} \int_0^{x/t} (1 - y^2/n)^{\frac{1}{2}(n-3)} dy \rightarrow \sigma^*(x/t) - \frac{1}{2}$$

holds uniformly for $\epsilon \leq t < \infty$. It follows that

$$(19) \quad \begin{aligned} A_{m_n} m_n^{-\frac{1}{2}} \int_{\epsilon}^{\infty} \int_0^{x/t} (1 - y^2/m_n)^{\frac{1}{2}(m_n-3)} dy d\rho_{m_n}(m_n^{\frac{1}{2}}t) \\ \rightarrow \int_{\epsilon}^{\infty} [\sigma^*(x/t) - \frac{1}{2}] d\tau(t) \end{aligned}$$

holds for every $x > 0$ and every $\epsilon > 0$ (such that $t = \epsilon$ is a continuity point of τ).

Furthermore, if $x > 0$ is fixed and $t \geq n^{-\frac{1}{2}}x$, then, by (7),

$$A_n n^{-\frac{1}{2}} \int_0^{x/t} (1 - y^2/n)^{\frac{1}{2}(n-3)} dy \leq A_n n^{-\frac{1}{2}} \int_0^{n^{\frac{1}{2}}} (1 - y^2/n)^{(n-3)} dy = \frac{1}{2};$$

so that

¹⁶ Cf., e. g., Wintner [22].

$$A_n n^{-\frac{1}{2}} \int_x^\epsilon \int_0^{x/t} (1 - y^2/n)^{\frac{1}{2}(n-3)} dy d\rho_n(n^{\frac{1}{2}}t) \leq \frac{1}{2}[\rho_n(n^{\frac{1}{2}}\epsilon) - \rho_n(x)].$$

Hence,

$$(20) \quad \limsup_{n \rightarrow \infty} A_{m_n} m_n^{-\frac{1}{2}} \int_{m_n^{-\frac{1}{2}}x}^\epsilon \int_0^{x/t} (1 - y^2/m_n)^{\frac{1}{2}(m_n-3)} dy d\rho_n(m_n^{\frac{1}{2}}t) \leq \frac{1}{2}[\tau(\epsilon) - \tau(+0)].$$

Finally, from (13),

$$(21) \quad \int_{+0}^\epsilon [\sigma^*(x/t) - \frac{1}{2}] d\tau(t) \leq \frac{1}{2} \int_{+0}^\epsilon d\tau(t) = \frac{1}{2}[\tau(\epsilon) - \tau(+0)].$$

The relations (14), (16), (19), (20), (21) obviously imply

$$(22) \quad \sigma(x) = \frac{1}{2} + \frac{1}{2}\tau(+0) + \int_{+0}^\infty [\sigma^*(x/t) - \frac{1}{2}] d\tau(t), \text{ if } x > 0;$$

while (11) is a consequence of (3 bis), (15) and (22). But (22) is equivalent to (12) in virtue of (11) and (4). This proves the second half of the statement italicized at the beginning of this section.

In order to prove the converse, notice first that if the distribution function $\tau(t)$ in (12) is the function $\tau^*(t)$ defined by

$$(23) \quad \tau^*(t) = \frac{1}{2} + \frac{1}{2} \operatorname{sgn}(t-1),$$

the corresponding distribution function σ in (12) is precisely σ^* . It is well known that σ^* is a distribution function of class Ω ; in fact, the function $\sigma = \sigma^*$ is known to belong, in virtue of (5_{n1}) and (8), to the function $\rho_n = \rho_n^* =$

$$(24) \quad \rho_n^*(r) = \int \prod_{k=1}^n (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}y_k^2} dy_k, \text{ where } \sum_{k=1}^n y_k^2 < r^2$$

(Gauss, Bravais, Maxwell; also Schoenberg¹⁷). Hence, it is seen¹⁸ that the function (12) is the 1-dimensional projection of the n -dimensional radially symmetric distribution function ϕ_n belonging to

¹⁷ Schoenberg [15], p. 817 (top).

¹⁸ This statement is an obvious consequence of (10) and the fact that if τ_1, τ_2, τ_3 are three distribution functions such that $\tau_1(0) = \tau_2(0) = \tau_3(0) = 0$, then

$$\int_0^\infty \tau_2(x/t) d\tau_3(t) = \int_0^\infty \tau_3(x/t) d\tau_2(t)$$

and

$$\int_0^\infty \tau_1(x/t) d_t \left[\int_0^\infty \tau_2(t/y) d\tau_3(y) \right] = \int_0^\infty \left[\int_0^\infty \tau_1(x/yt) d\tau_3(y) \right] d\tau_2(t).$$

The first of these relations is merely an integration by parts; the second clearly is true if τ_2 is a step-function, so that the relation holds in general, in view of the definition of Stieltjes integrals.

$$(25) \quad \rho_n(r) = \int_0^\infty \rho_n^*(r/t) d\tau(t)$$

in virtue of (2_n). This completes the proof of the italicized statement.

3. In the sequel, it will be necessary to make use of the fact that if σ is a distribution function of class Ω , the function τ occurring in (12) is unique. This is easily proved by using Fourier-Stieltjes transforms; cf. the remark in § 1 concerning ρ_k and ρ_n . Incidentally, the uniqueness of τ , when combined with standard application of the theorem of Helly, obviously implies that (15) is valid without applying any selection, i. e., by placing $m_n = n$.

The problem of replacing the sheaf of normal distribution functions $\sigma^*(x/t)$ in the representation (12) of a distribution function of class Ω by a sheaf of arbitrary distribution functions $\omega(x, y)$ of class Ω will now be considered. Let $\tau(t, y)$ denote a function which is defined for $0 \leq t < \infty$, $0 \leq y < \infty$ in such a way that, for every fixed $t \geq 0$, $\tau(t, y)$ is a Baire function of y , $0 \leq y < \infty$; and that it is, for every fixed $y \geq 0$, a distribution function, i. e., a non-decreasing function satisfying the boundary conditions $\tau(0, y) = 0$, $\tau(+\infty, y) = 1$. Let $\omega(x, y)$ denote, for a fixed y , the distribution function of class Ω corresponding to $\tau(t, y)$ in virtue of (12), so that

$$(26) \quad \omega(x, y) = \int_0^\infty \sigma^*(x/t) d\tau(t, y).$$

It is clear that $\omega(x, y)$ is, for a fixed x , a Baire function of y (≥ 0). As above, it can easily be shown that if $\xi(t)$, $-\infty < t < +\infty$, is a distribution function satisfying $\xi(0) = 0$, then the distribution function

$$(27) \quad \sigma(x) = \int_0^\infty \omega(x, t) d\xi(t)$$

is a distribution function of class Ω . On the other hand, it will be proved that if $\sigma(x)$ is a distribution function of class Ω associated with the function $\tau(t)$ in virtue of (12), then $\sigma(x)$ has a representation of the form (27) if and only if there exists a distribution function $\xi(t)$ such that $\xi(0) = 0$, $\xi(+\infty) = 1$ and

$$(28) \quad \tau(t) = \int_0^\infty \tau(t, y) d\xi(y).$$

Suppose first that $\sigma(x)$ has a representation of the form (27). Define a sequence of distribution functions $\tau^m(t)$, which tend to $\xi(t)$ as $m \rightarrow \infty$ and which are of the form

$$\tau^m(t) = \sum_{i=1}^m a_{im} \tau^*(t/h_{im}), \quad (a_{im} > 0, h_{im} > 0, \sum_{i=1}^m a_{im} = 1),$$

where τ^* is defined in (23); so that by (27) and the definition of Stieltjes' integrals,

$$\begin{aligned}
 (29) \quad \sigma(x) &= \lim_{m \rightarrow \infty} \int_0^\infty \omega(x, t) d\tau^m(t) = \lim_{m \rightarrow \infty} \sum_{i=1}^m a_{im} \omega(x, h_{im}) \\
 &= \lim_{m \rightarrow \infty} \sum_{i=1}^n a_{im} \int_0^\infty \sigma^*(x/t) d_i \tau(t, h_{im}) \\
 &= \lim_{m \rightarrow \infty} \int_0^\infty \sigma^*(x/t) d_i \left[\int_0^\infty \tau(t, y) d\tau^m(y) \right] \\
 &= \int_0^\infty \sigma^*(x/t) d_i \left[\int_0^\infty \tau(t, y) d\xi(y) \right].
 \end{aligned}$$

Hence, (28) follows from (12) in virtue of the uniqueness of the distribution function τ in (12). And also the converse of the italicized statement follows from (29), since the preceding steps are obviously reversible.

Suppose that the sheaf of distribution functions $\omega(x, y)$ has the property that, for some fixed $L > 0$, the function $\sigma^*(x/L)$ may be represented in the form (27); so that there exists a distribution function $\xi_L(y)$, such that $\xi_L(0) = 0$, $\xi_L(+\infty) = 1$, and

$$(30) \quad \sigma^*(x/L) = \int_0^\infty \omega(x, t) d\xi_L(t).$$

Then, by the italicized statement just proved, the function $\tau = \tau^*(t/L)$, which corresponds to (30) in virtue of (12), satisfies

$$(31) \quad \tau^*(t/L) = \int_0^\infty \tau(t, y) d\xi_L(y).$$

Let $y = T^L$ denote an arbitrary point in the spectrum¹⁹ of $\xi_L(y)$, and let $t < L$, $\epsilon > 0$. Then, by (31), (23),

$$\begin{aligned}
 0 = \tau^*(t/L) &\geq \int_{T^L - \epsilon}^{T^L + \epsilon} \tau(t, y) d\xi_L(y) \\
 &\geq [\xi_L(T^L + \epsilon) - \xi_L(T^L - \epsilon)] \inf_{T^L - \epsilon \leq y \leq T^L + \epsilon} \tau(t, y).
 \end{aligned}$$

Hence,

$$\liminf_{y \rightarrow T^L} \tau(t, y) = 0 \text{ if } t < L.$$

It follows that there exist a distribution function $\tau_1(t)$, $-\infty < t < +\infty$, and a sequence of positive numbers T_n such that $T_n \rightarrow T^L$ and such that

$$\tau(t, T_n) \rightarrow \tau_1(t), \text{ as } n \rightarrow \infty; \text{ finally, } \tau_1(t) = 0 \text{ if } t < L.$$

Thus, by the term-by-term integration theorem of Helly,

¹⁹ A point is said to belong to the spectrum of a function if the function is not constant in any interval containing this point in its interior.

$$\lim_{n \rightarrow \infty} \omega(x, T_n) = \int_{L-0}^{\infty} \sigma^*(x/t) d\tau_1(t), \quad (T_n \rightarrow T^L).$$

It is similarly shown that there exist a sequence of positive numbers T'_n such that $T'_n \rightarrow T^L$, and a distribution function $\tau_2(t)$ such that $\tau_2(t) = 1$, if $t > L$ and $\tau(t, T'_n) \rightarrow \tau_2(t)$ as $n \rightarrow \infty$; so that

$$\lim_{n \rightarrow \infty} \omega(x, T'_n) = \int_0^{L+0} \omega(x/t) d\tau_2(t), \quad (T'_n \rightarrow T^L).$$

It follows that if the distribution function $\omega(x, y)$ tends to $\omega(x, y_0)$ as $y \rightarrow y_0$ (in the sense of monotone functions) for every y_0 , $0 \leq y_0 < \infty$, then every distribution function of class Ω may be represented in the form (27) if and only if there exists for every $L > 0$ at least one $T = T^L$ such that $\omega(x, T) = \sigma^*(x/L)$, $-\infty < x < \infty$.

Suppose, in particular, that the sheaf of distribution functions $\omega(x, t)$ is of the form $\omega(x, t) = \omega(x/t)$, where $\omega(x)$ is an arbitrary distribution function of class Ω ; so that, by § 2,

$$\omega(x) = \int_0^{\infty} \sigma^*(x/t) d\tau_{\omega}(t)$$

holds for a suitable distribution function $\tau = \tau_{\omega}$ satisfying (11). A distribution function σ which may be represented by means of ω in the form

$$\sigma(x) = \int_0^{\infty} \omega(x/t) d\xi(t), \quad x \neq 0,$$

where $\xi(t)$ is a distribution function satisfying $\xi(0) = 0$, will be said to be of class $\Omega(\omega)$. In this particular case, the preceding results are seen to be to the effect that a distribution function (12) is of class $\Omega(\omega)$ if and only if there exists a distribution function $\xi(t)$ which vanishes at $t = 0$ and satisfies

$$(32) \quad \tau(t) = \int_0^{\infty} \tau_{\omega}(t/y) d\xi(y);$$

furthermore, every distribution function of class Ω is of class $\Omega(\omega)$ if and only if there exists a positive number T^{ω} such that $\omega(x/T^{\omega}) = \sigma^*(x)$, $-\infty < x < +\infty$. (The italicized statement of § 2 is the particular case $T^{\omega} = 1$).

If, in addition, use is made of the Stieltjes-Fubini relation which is the second formula of footnote 18, one sees that if the distribution function $\sigma(x)$ is of class $\Omega(\omega)$ and if the distribution function $\mu(x)$ is of class $\Omega(\sigma)$, then $\mu(x)$ is of class $\Omega(\omega)$.

4. As an application of these statements, consider the symmetric stable distribution functions; that is, the distribution functions whose Fourier-Stieltjes transforms are $\exp(-|u|^{\gamma})$, $0 < \gamma \leq 2$ (the distribution function,

whose Fourier-Stieltjes is identically 1, also is symmetric and stable, but will be excluded as trivial). It is known²⁰ that these distribution functions are of class Ω . Let $\sigma_\gamma(x)$ be the distribution function whose Fourier-Stieltjes transform is $\exp(-|u|^\gamma)$, $0 < \gamma \leq 2$. Thus, there exists a distribution function $\tau^\gamma(t)$ such that $\tau^\gamma(0) = 0$ and

$$(33) \quad \sigma_\gamma(x) = \int_0^\infty \sigma^*(x/t) d\tau^\gamma(t), \text{ where } \sigma^* = \sigma_2,$$

if σ^* denotes the same distribution function as in (13), except that the unit of x is different,

$$(13 \text{ bis}) \quad \sigma_2(x) = (2\pi)^{-1} \int_{-\infty}^{2^{-1/2}x} e^{-1/2 y^2} dy.$$

The relation (33) implies that

$$(34) \quad \exp(-|u|^\gamma) = \int_0^\infty \exp(-|ut|^2) d\tau^\gamma(t), \quad -\infty < u < +\infty.$$

This merely states that the Fourier-Stieltjes transform of the function on the left of (33) is the same as the Fourier-Stieltjes transform of the function on the right.

On replacing $|u|$ by $|u|^{\beta/\gamma}$, $0 < \beta \leq \gamma \leq 2$, and changing the integration variable from t to $t^{\beta/\gamma}$, one can write (34) in the form

$$(35) \quad \exp(-|u|^\beta) = \int_0^\infty \exp(-|ut|^{2\beta/\gamma}) d\tau^\gamma(t^{\beta/\gamma}), \quad -\infty < u < +\infty,$$

or

$$(36) \quad \sigma_\beta(x) = \int_0^\infty \sigma_{2\beta/\gamma}(x/t) d\tau^\gamma(t^{\beta/\gamma}), \quad 0 < \beta \leq \gamma \leq 2.$$

Since β, γ are arbitrary ($0 < \beta \leq \gamma \leq 2$), this relation is equivalent to the first half of the statement: $\sigma_\alpha(x)$ is of class $\Omega(\sigma_\beta)$ if and only if $\alpha \leq \beta$.

To prove the second half of this statement, suppose that σ_α is of class $\Omega(\sigma_\beta)$, $0 < \beta < \alpha \leq 2$; so that there exists a distribution function $\tau_{\alpha\beta}(t)$ which vanishes for $t = 0$ and satisfies

²⁰ Wintner [21]. This result was rediscovered by Schoenberg [16], pp. 532-533 (cf. Blumenthal [11]), who used methods equivalent (cf. Haviland [8], II, p. 382) to those applied *loc. cit.* [20], where the proof, in fact, was based on the multidimensional analogue of Lévy's continuity theorem (cf. Haviland [8], II). Incidentally, cf. Wiener and Wintner [19], pp. 241-242.

It may be mentioned in this connection that Theorem 3 of Schoenberg [16] is merely a corollary of the classical representation of the infinitely divisible laws which is due P. Lévy (who, in fact, does not assume the symmetry of the distributions).

$$(37) \quad \sigma_a(x) = \int_0^\infty \sigma_\beta(x/t) d\tau_{a\beta}(t).$$

Then, by the same reasoning which deduced (36) from (33),

$$(38) \quad \sigma_2(x) \equiv \sigma^*(x) = \int_0^\infty \sigma_{2\beta/a}(x/t) d\tau_{a\beta}(t^{2/a}).$$

This would imply that there exists a positive number $T = T(\sigma_{2\beta/a})$ such that $\sigma_{2\beta/a}(x/T) \equiv \sigma_2(x)$, or $\exp(-|uT|^{2\beta/a}) \equiv \exp(-|u|^2)$, $-\infty < u < +\infty$. This contradiction establishes the theorem.

The theorem just proved and the last italicized theorem of § 3 imply that if $\alpha < \beta$, the class $\Omega(\sigma_\alpha)$ is a proper subset of class $\Omega(\sigma_\beta)$.

5. The standard methods described at the beginning of the Introduction represent asymptotic approaches to the normal distributions. Another approach to these distributions is connected with the well-known fact that the stable distribution σ_γ has a finite standard deviation only in the normal case $\gamma = 2$. In what follows, there will be considered still another approach to the normal distributions of radial symmetry. This approach might be of interest in view of Maxwell's deduction of his distribution law of velocities.

Let ϕ_n be an arbitrary radially symmetric distribution function on the Euclidean space R_n . Let ϕ_k be the k -dimensional projection of ϕ_n , finally ρ_n, ρ_k the corresponding functions $(2_n), (2_k)$. Since the function (9) is absolutely continuous, it follows from (10) that if $k < n$, then ρ_n is absolutely continuous on the open half-line $(0, +\infty)$; so that

$$(39) \quad \rho_k(x) = \rho_k(+0) + \int_0^x (d\rho_k(r)/dr) dr, \quad x > 0; \quad (k = 1, \dots, n-1).$$

Hence, the k -dimensional distribution function ϕ_k , where $k = 1, \dots, n-1$ may be decomposed into a linear combination of two k -dimensional distribution functions ϕ_k^I, ϕ_k^{II} of radial symmetry,

$$(40) \quad \phi_k = \lambda \phi_k^I + (1-\lambda) \phi_k^{II}, \quad 0 \leq \lambda \leq 1; \quad (k = 1, \dots, n-1),$$

where $\phi_k^I(E_k^0) = 1$ if E_k^0 denotes the Borel set consisting of the single point which is the origin of R_k , and ϕ_k^{II} is absolutely continuous; so that there exists a non-negative function

$$\delta_k = \delta_k(x_1, \dots, x_k) = \delta_k(|x_1|^2 + \dots + |x_k|^2)^{\frac{1}{2}}$$

of the position (x_1, \dots, x_k) in R_k for which

$$(41) \quad \phi_k^{II}(E_k) = \int_{E_k} \delta_k(|x_1|^2 + \dots + |x_k|^2)^{\frac{1}{2}} dx_1 \dots dx_k.$$

It follows from (2_k) , (39) and (40) that $\lambda = \rho_k(+0) = \rho_n(+0)$ and

$$(42) \quad \frac{(2\pi)^{\frac{1}{2}k}}{\Gamma(\frac{1}{2}k)} \delta_k(r) = r^{-k+1} \frac{d\rho_k(r)}{dr}, \quad (\text{for almost all } r > 0),$$

$(2\pi)^{\frac{1}{2}k}/\Gamma(\frac{1}{2}k)$ being the Euclidean measure of the boundary of the k -sphere of radius 1.

Define, for $0 < r < \infty$, a non-decreasing function, $v_k(r)$, $k = 1, \dots, n$, by placing

$$(43) \quad \frac{(2\pi)^{\frac{1}{2}k}}{\Gamma(\frac{1}{2}k)} v_k(r) = 1 - \int_r^\infty x^{-k+1} d\rho_k(x), \quad r > 0.$$

Thus, for $k = 1, \dots, n$,

$$(44) \quad \rho_k(r) = 1 - \frac{(2\pi)^{\frac{1}{2}k}}{\Gamma(\frac{1}{2}k)} \int_r^\infty x^{k-1} dv_k(x), \quad r > 0;$$

also, for $k = 1, \dots, n-1$,

$$(45) \quad \delta_k(r) = dv_k(r)/dr, \quad (\text{for almost all } r > 0)$$

The relation (45) is meaningless for $k = n$, unless the arbitrary function ϕ_n has a decomposition similar to (40), implying the existence of a δ_n .

It will be shown that if ϕ_n is an n -dimensional radially symmetric distribution function such that for some fixed k ($0 < k < n$), and for some pair of positive constants c, C , one has

$$(46) \quad v_n(cr) = C^k v_{n-k}(r), \quad r > 0,$$

then ϕ_n is a radially symmetric normal distribution except for a possible jump at the origin. This means that ϕ_n may be written in terms of a non-negative constant $\lambda \leq 1$ in the form

$$(47) \quad \phi_n = \lambda \phi_n^I + (1 - \lambda) \phi_n^{II}, \quad 0 \leq \lambda \leq 1,$$

where $\phi_n^I(E_n^0) = 1$, and ϕ_n^{II} is a radially symmetric n -dimensional normal distribution.

In order to prove this theorem, note that, in view of (43) and (46), the absolute continuity of ρ_{n-k} for $r > 0$ implies the absolute continuity of v_{n-k}, v_n, ρ_n for $r > 0$. Thus, under these conditions, equations similar to (40), (41), and (45) hold for $k = n$. Since ϕ_{n-k}^{II} is the projection of ϕ_n^{II} , it is clear that

$$(48) \quad \delta_{n-k}(r) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \delta_n(|r^2 + x_1^2 + \dots + x_k^2|) dx_1 \dots dx_k;$$

so that

$$(49) \quad c\delta(cr) = C^k \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \delta(|r^2 + x_1^2 + \dots + x_k^2|) dx_1 \dots dx_k,$$

if $\delta(r)$ denotes the common value of $\delta_n(r)$ and $c^{-1}C^k\delta_{n-k}(r/c)$; cf. (45) and (46). Since δ is a density, repeated integration of (49) shows that

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \delta(|r^2 + x_1^2 + \dots + x_m^2|) dx_1 \dots dx_m < \infty, \quad (\delta \geq 0)$$

for every positive integer m . It follows, therefore, from (48) and (49) that, up to a factor of proportionality depending on m , the functions $C^k \delta(r)$, $c \delta(cr)$ are the densities of an m -dimensional radially symmetric distribution function and its $(m - k)$ -dimensional projection, respectively (unless $\delta(r) = 0$ for every $r > 0$; but this trivial case may be discarded, for in this case the statement (49) is satisfied by $\lambda = 1$). Thus, one can introduce a 1-dimensional distribution function by placing

$$(50) \quad \sigma(x) = \int_{-\infty}^x \delta(|r|) dr / \int_{-\infty}^{+\infty} \delta(|y|) dy, \quad -\infty < x < +\infty.$$

Clearly, this $\sigma(x)$ is, for $m = 1, 2, \dots$, the projection of an $(mk + 1)$ -dimensional radially symmetric distribution whose density is proportional to $\delta(r)$. Consequently, by (12),

$$(51) \quad \sigma(x) = \int_0^{\infty} \sigma^*(x/t) d\tau(t), \quad -\infty < x < +\infty,$$

where $\tau(t)$ is a distribution function satisfying (11). Since (50), (51) and (13) imply that

$$(52) \quad \delta(|r|) / \int_{-\infty}^{+\infty} \delta(|y|) dy = (2\pi)^{-\frac{1}{2}} \int_0^{\infty} t^{-1} \exp(-\frac{1}{2}r^2/t^2) d\tau(t),$$

it follows from (49) that

$$c \delta(c|r|) / \int_{-\infty}^{+\infty} \delta(|y|) dy = C^k (2\pi)^{-\frac{1}{2}(1+k)} \int_0^{\infty} t^{k-1} \exp(-\frac{1}{2}r^2/t^2) d\tau(t).$$

On integrating this relation between $r = -\infty$ and $r = x$, one sees from (50) and (13) that

$$\sigma(cx) = C^k (2\pi)^{-\frac{1}{2}k} \int_0^{\infty} \sigma^*(x/t) t^k d\tau(t).$$

It follows, therefore, by comparison with (51) that

$$(53) \quad \tau(cy) = C^k (2\pi)^{-\frac{1}{2}k} \int_0^y t^k d\tau(t), \quad 0 \leq y < \infty.$$

But it is clear that (53) cannot hold unless $c = 1$; in which case

$$\tau(t) = \tau^*((2\pi)^{-\frac{1}{2}} Ct),$$

where τ^* is defined as in (23). Hence, from (52),

$$\delta(|r|) / \int_{-\infty}^{+\infty} \delta(|y|) dy = \frac{1}{2\pi} C \exp(-C^2 r^2 / 4\pi).$$

This completes the proof of the last italicized statement.

6. It is clear from the proof and from the Helly theory of monotone functions that the theorem just proved may be generalized as follows: *Let*

$\sigma(x)$ be the 1-dimensional projection of an n -dimensional radially symmetric distribution function ϕ_n , $n = 1, 2, \dots$. Let $v_n(r)$ be the function (43) associated with ϕ_n , and suppose that there exists a non-decreasing function $v(r)$ which is the limit of the sequence of functions

$$\{[v_n(r) - v_n(+0)]/[1 - v_n(+0)]\}$$

in the sense of the theory of monotone functions. Then there exist two constants c, C ($0 < C, 0 \leq c \leq 2C/\pi^{\frac{1}{2}}$) such that

$$v(r) = c \int_0^r \exp(-C^2 x^2) dx.$$

It is also clear from the above proof that an absolutely integrable solution $\delta(r)$ of the equation (49), i. e., of the Abel integral equation

$$c\delta(cr) = C^k \int_r^{+\infty} \delta(x) (x^2 - r^2)^{\frac{1}{2}(k-2)} x dx,$$

exists only if $c = 1$; in which case it is proportional to $\exp(-C^2 r^2/4\pi)$.

7. The italicized statement of § 5 implies a characterization of the n -dimensional distribution functions which are product distributions with respect to every coördinate system.

In § 1, the k -dimensional projection of an n -dimensional radially symmetric distribution function was defined by considering a k -dimensional hyperplane R_k through the origin of the n -dimensional Euclidean space R_n . Because of the radial symmetry, the projection was independent of the choice of the hyperplane R_k . It is clear that if one considers an arbitrary n -dimensional distribution function $\psi_n(E_n)$ (not necessarily of radial symmetry), one can obtain a sheaf of k -dimensional projections $\psi_k(E_k; R_k)$, where the argument of the distribution function ψ_k is a k -dimensional Borel set E_k on the hyperplane R_k on which $\psi_n(E_n)$ is projected. An n -dimensional distribution function $\psi_n(E_n)$ is said to be a product distribution²¹ if there exists a positive integer $k < n$, a k -dimensional hyperplane R_k and its $(n-k)$ -dimensional normal hyperplane R_{n-k} such that if $E_n = E_k \times E_{n-k}$ is an n -dimensional Borel set whose projections on R_k, R_{n-k} are E_k, E_{n-k} , respectively, then

$$(54) \quad \psi_n(E_k \times E_{n-k}) = \psi_k(E_k; R_k) \psi_{n-k}(E_{n-k}; R_{n-k}).$$

The Fourier-Stieltjes criterion for ψ_n to be a product distribution is that there

²¹ This is slightly more general than the usual concept of a product distribution, in which (54) is replaced by

$$\psi_n(E_1^1 \times \dots \times E_1^n) = \psi_1(E_1^1; R_1) \dots \psi_1(E_1^n; R_{n_1}),$$

where $R_1^1, \dots, R_{n_1}^1$ are n mutually perpendicular lines. A distribution which is a product distribution in this sense is clearly a product distribution in the sense of (54), but not conversely.

exists at least one rectangular coordinate system in R_n with reference to which the Fourier-Stieltjes transform $\Lambda(u_1, \dots, u_n)$ of ψ_n can be written as the product

$$(55) \quad \Lambda(u_1, u_2, \dots, u_n) = \Lambda(u_1, \dots, u_k, 0, \dots, 0) \Lambda(0, \dots, 0, u_{k+1}, \dots, u_n).$$

(In this coordinate system, the hyperplane R_k in (54) is defined by the equations $x_{k+1} = 0, \dots, x_n = 0$).

It will now be proved²² that if $\psi_n(E_n)$, where $n \geq 2$, is an n -dimensional distribution function, then, for a fixed k , (54) holds for every pair of orthogonal hyperplanes R_k, R_{n-k} if and only if either $\psi_n(E_n^0) = 1$ or there exist constants $a (> 0)$, b_1, \dots, b_n such that

$$(56) \quad \psi_n(E_n) = (a\pi^{-\frac{1}{2}})^n \int_{E_n} \exp[-a^2 \sum_{j=1}^n (x - b_j)^2] dx_1 \cdots dx_n.$$

It is understood that E_n^0 denotes the Borel set consisting of one point in R_n (not necessarily the origin).

The first half of the theorem is trivial. In order to prove its second half, let $P = (u_1, \dots, u_n)$ be a point in the space of the Fourier-Stieltjes transform $\Lambda(u_1, \dots, u_n)$. Then the assumptions of the theorem imply that

$$(57) \quad \Lambda(P) = \Lambda(P^k) \Lambda(P^{n-k}),$$

where P^k, P^{n-k} are the projections of P on an arbitrary pair of orthogonal k - and $(n-k)$ -dimensional hyperplanes through the origin of (u_1, \dots, u_n) -space, respectively.

Suppose first that the distribution function ψ_n is symmetric with respect to the origin of R_n , i. e., $\Lambda(u_1, \dots, u_n) = \Lambda(-u_1, \dots, -u_n)$. It will be shown that ψ_n is then of radial symmetry. In fact, let P, Q be two distinct points on any sphere with its center at the origin O of the (u_1, \dots, u_n) -space. Consider the plane POQ , the pair of lines which bisect the angles formed by the lines OP and OQ , and a pair of orthogonal hyperplanes containing these lines and having the dimension numbers k and $n-k$, respectively. Then

$$\Lambda(P) = \Lambda(P^k) \Lambda(P^{n-k}) \quad \text{and} \quad \Lambda(Q) = \Lambda(Q^k) \Lambda(Q^{n-k}),$$

where $P^k, P^{n-k}, Q^k, Q^{n-k}$ are the projections of P and Q on these hyperplanes, respectively. It is clear that if the points P^k, Q^k do not coincide, then they

²² This problem was considered by Maria-Pia Geppert, "Una propriet  caratteristica della distribuzione de Bravais," *Giornale dell' Istituto Italiano degli Attuari*, vol. 7 (1936), pp. 378-391. Her considerations were recently rediscovered by M. Kac [10]. Actually, the final result of Kac is incorrect, since his conclusion is that either ψ_n has a jump of 1 at the origin or (56) must hold with $b_1 = b_2 = \dots = b_n = 0$. In the case of polar symmetry, Kac used Cauchy's functional equation, which will now be avoided by applying the theorem of § 5.

are situated symmetrically with respect to the origin O ; the same holds for the pair P^{n-k}, Q^{n-k} . In virtue of the polar symmetry of ψ_n , it follows that $\Lambda(P) = \Lambda(Q)$, which establishes the radial symmetry of ψ_n .

Since $\psi_k(E_k; R_k)$, $\psi_{n-k}(E_{n-k}; R_{n-k})$ are projections of the radially symmetric distribution ψ_n , they are absolutely continuous on R_k, R_{n-k} , respectively, if the origin is removed. It follows, therefore, from (54) that ψ_n is absolutely continuous on R_n with the origin removed, and that the density $\delta_n(x_1, \dots, x_n)$ of ψ_n then is the product of the densities $\delta_k(x_1, \dots, x_k)$, $\delta_{n-k}(x_{k+1}, \dots, x_n)$ of ψ_k and ψ_{n-k} . Consequently, the radial symmetry of ψ_n implies that there exist a function $\delta(x)$, $-\infty < x < +\infty$, and two positive constants c_1, c_2 such that

$$\delta(|x_1^2 + \dots + x_n^2|^{\frac{1}{2}}) = \delta_n(x_1, \dots, x_n), \delta(|x_1^2 + \dots + x_k^2|^{\frac{1}{2}}) = c_1 \delta_k(x_1, \dots, x_k)$$

and

$$\delta(|x_{k+1}^2 + \dots + x_n^2|^{\frac{1}{2}}) = c_2 \delta_{n-k}(x_{k+1}, \dots, x_n).$$

Thus, it is easy to see that the assumptions of the theorem of § 5 are satisfied; so that ψ_n is a radially symmetric normal distribution except for a possible jump at the origin. However, the product condition implies that the jump at the origin is either 0 or 1. This concludes the proof of the last italicized statement in case ψ_n is symmetric with respect to the origin of R_n .

In order to complete the proof, consider the n -dimensional distribution function $\phi_n(E_n)$ whose Fourier-Stieltjes transform is

$$\Lambda(u_1, \dots, u_n) \Lambda(-u_1, \dots, -u_n).$$

Then $\phi_n(E_n)$ is symmetric with respect to the origin and satisfies the conditions of the theorem. Hence, $\phi_n(E_n)$ either is a radially symmetric normal distribution or ϕ_n has a jump of 1 at the origin. Thus,²³

$$\Lambda(u_1, \dots, u_n) \Lambda(-u_1, \dots, -u_n) = \exp[-a^2(u_1^2 + \dots + u_n^2)], \\ 0 \leq a < \infty.$$

It follows that Λ does not vanish for any (u_1, \dots, u_n) ; so that

$$(58) \quad \Lambda(u_1, \dots, u_n) = \exp[-a^2(u_1^2 + \dots + u_n^2) + g(u_1, \dots, u_n)]$$

holds for a suitable continuous function $g(u_1, \dots, u_n)$ which satisfies the condition

$$g(u_1, \dots, u_n) = -g(-u_1, \dots, -u_n).$$

²³ The balance of the proof could be based (cf. Kac [10]) on an application of a theorem formulated as a conjecture by Lévy and subsequently proved by Cramér [6]. But this rather deep theorem, for which only a complex function-theoretical proof is available today, may be avoided in this case.

It follows from (57) and (58), by choosing

$$P = (u_1, \dots, u_n), \quad P^k = (u_1, 0, \dots, 0), \quad P^{n-k} = (0, u_2, \dots, u_n),$$

that $g(u_1, \dots, u_n) = g(u_1, 0, \dots, 0) + g(0, u_2, \dots, u_n)$, i. e.,

$$(59) \quad \begin{aligned} g(u_1, \dots, u_n) &= g_1(u_1) + \dots + g_n(u_n), \quad \text{where} \\ g_i(u_i) &= g(0, \dots, 0, u_i, 0, \dots, 0) \end{aligned}$$

is a continuous odd function of u_i .

On applying (57), (58) and (59) to $P \equiv (u^2 + v^2, 0, \dots, 0)$,

$$P^k \equiv (u^2, uv, 0, \dots, 0), \quad P^{n-k} \equiv (v^2, -uv, 0, \dots, 0),$$

one obtains

$$g_1(u^2 + v^2) = g_1(u^2) + g_1(v^2),$$

if use is made of the fact that g_2 is odd and $g_i(0) = 0$. This implies that there exists a constant c_1 such that $g_1(u^2) = c_1 u^2$; so that, since g_1 is odd, $g_1(u) = c_1 u$. Similarly, $g_j(u) = c_j u$ for $j = 2, \dots, n$. Hence, (58) reduces, in view of (59), to

$$\Lambda(u_1, \dots, u_n) = \exp\left[-\sum_{j=1}^n (a^2 u_j^2 - c_j u_j)\right].$$

But since Λ is a Fourier-Stieltjes transform of a distribution function, $|\Lambda| \leq 1$, so that the constants c_j are purely imaginary, i. e., $c_j = ib_j$ and

$$(60) \quad \Lambda(u_1, \dots, u_n) = \exp\left[-\sum_{j=1}^n (a^2 u_j^2 - ib_j u_j)\right], \quad 0 \leq a < \infty.$$

Since (60) is known to be the Fourier-Stieltjes transform of an n -dimensional distribution of the particular type mentioned in the theorem, the proof is complete.

8. For a positive number p which need not be an integer, let S_n^p be the solid characterized by the inequality

$$(61) \quad S_n^p : \sum_{j=1}^n |x_j|^p \leq 1$$

in the Euclidean space $R_n: (x_1, \dots, x_n)$. It will be shown that if $\lambda_n^p(x)$, $-\infty < x < +\infty$, denotes the one-dimensional distribution function which one obtains by projecting on a coördinate axis of R_n the n -dimensional equi-distribution on S_n^p , the density of probability of λ_n^p is

$$(62) \quad \frac{d}{dx} \lambda_n^p(x) = \begin{cases} \text{const. } (1 - |x|^p)^{(n-1)/p}, & \text{if } |x| < 1, \text{ where} \\ \text{const.} = \frac{\frac{1}{2} p \Gamma(1 + n/p)}{\Gamma\left(\frac{1}{p}\right) \Gamma\left(\frac{n-1}{p} + 1\right)}; & \\ 0, & \text{if } |x| > 1. \end{cases}$$

In order to prove (62), let n and p be fixed, and let s denote a continuous parameter which varies between 0 and 1. A straightforward homogeneity consideration shows that the n -dimensional volume of that infinitesimal portion of the solid (61) which lies between the two hyperplanes $x_1 = s$, $x_1 = s + ds$ is proportional to $(1 - s^p)^q ds$, where $q = (n-1)/p$. Since the whole solid (61) is contained between the two hyperplanes $x_1 = \pm 1$, and is symmetric with respect to the hyperplane $x_1 = 0$, it follows that (62) holds for some const. > 0 . Finally, the value of this constant is obvious from

$$\int_{-1}^1 (1 - |x|^p)^{(n-1)/p} dx = \frac{2}{p} \int_0^1 (1 - y)^{(n-1)/p} y^{1/p-1} dy = \frac{2}{p} B\left(\frac{n-1}{p} + 1, \frac{1}{p}\right),$$

and from the fact that the total probability represented by $\lambda_n^p(x)$, $-\infty < x < +\infty$, is unity. This completes the proof of (62).

A corollary of (62) is that if $S_n^p(r)$ denotes, for a fixed $r > 0$, the solid which one obtains by writing r^p instead of 1 on the right of the inequality (61), the projection on a coordinate axis of the equidistribution on $S_n^p(n^{1/p})$ tends, as $n \rightarrow \infty$, to the distribution function which has a density proportional to $\exp(-|x|^p/p)$ for $-\infty < x < +\infty$.

In fact, if p is fixed and $n \rightarrow \infty$, then

$$\Gamma\left(\frac{n}{p} + 1\right) / \Gamma\left(\frac{n-1}{p} + 1\right) \sim \frac{n^{1/p}}{p^{1/p}}$$

(Stirling); while

$$\lim_{n \rightarrow \infty} \left(1 - \left|\frac{x}{n^{1/p}}\right|^p\right)^{(n-1)/p} = \exp\left(-\frac{|x|^p}{p}\right) \quad \text{for } -\infty < x < +\infty.$$

Hence, from (62),

$$(63) \quad \lim_{n \rightarrow \infty} \frac{d}{dx} \frac{1}{n^p} \lambda_n^p\left(\frac{x}{n^{1/p}}\right) = \frac{\exp(-|x|^p/p)}{2p^{1/p}\Gamma(1 + p^{-1})}; \quad -\infty < x < +\infty.$$

But it is clear for reasons of homogeneity that, if $r > 0$, the distribution function $r^{-1}\lambda_n^p(x/r)$, $-\infty < x < +\infty$, belongs to $S_n^p(r)$ in the same way as $\lambda_n^p(x)$ belongs to $S_n^p = S_n^p(1)$. Hence, (63) is equivalent to the last italicized statement.

Remark. If $L_n^p(u)$, $-\infty < u < +\infty$, denotes the Fourier transform of $\lambda_n^p(x)$, $-\infty < x < +\infty$, then, according to (62),

$$(64) \quad L_n^p(u) = \frac{p\Gamma(1 + n/p)}{\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{n-1}{p} + 1\right)} \int_0^1 (1 - x^p)^{(n-1)/p} \cos(ux) dx.$$

It is seen from the integral definition of the Bessel functions $J_\nu(u)$, that (64) reduces for $p = 2$ to

$$(65) \quad L_n^2(u) = J_{\frac{1}{2}n}^*(u)/J_{\frac{1}{2}n}^*(0),$$

if $J_\nu^*(u)$ denotes $J_\nu(|u|)/|u|^\nu$. On the other hand, if $\bar{L}_n^2(u)$ denotes the Fourier transform of the distribution function $\bar{\lambda}_n^2(x)$, $-\infty < x < +\infty$, which belongs to the equidistribution on the boundary $\bar{S}_n^2: \sum_{j=1}^n x_j^2 = 1$ of $S_n^2: \sum_{j=1}^n x_j^2 \leq 1$ in the same way as $\lambda_n^2(x)$ belongs to S_n^2 itself, then, as mentioned in the Introduction, it is well known that

$$(66) \quad \bar{L}_n^2(u) = J_{\frac{1}{2}n-1}^*(u)/J_{\frac{1}{2}n-1}^*(0).$$

Since comparison of (65) and (66) shows that $\bar{L}_{n+2}^2(u) = L_n^2(u)$, it follows that

$$(67) \quad \bar{\lambda}_{n+2}^2(x) = \lambda_n^2(x), \quad -\infty < x < +\infty; \quad (n = 1, 2, \dots).$$

In other words, the distribution which is the projection on a diameter of the equidistribution on the interior of the n -dimensional unit sphere is identical with the distribution which is the projection on a diameter of the equidistribution on the boundary of the $(n+2)$ -dimensional unit sphere.²⁴ (Needless to say, this fact may be verified also by calculating the volumes of the spherical segments involved.) Actually, the explicit relation (67) may be interpreted as an essential refinement of a known phenomenon in functional analysis;²⁵ that is, of the fact that, as $n \rightarrow \infty$ an overwhelming portion of the sphere $x_1^2 + \dots + x_n^2 < 1$ concentrates on its boundary $x_1^2 + \dots + x_n^2 = 1$.

QUEENS COLLEGE,
THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

- [1] Blumenthal, L. M., "Distance geometries," *University of Missouri Studies*, vol. 13 (1938), no. 2.
- [2] Boltzmann, L., *Vorlesungen über Gastheorie*, vol. 1 (1896); vol. 2 (1898); Leipzig.
- [3] Borel, E., *Mécanique statistique classique*, (Paris), Gauthier-Villars, 1925.
- [4] Borel, E., *Introduction géométrique à quelques théories physiques* (Paris), Gauthier-Villars, 1914.

²⁴ Cf. Borel [4]; Lévy [12], [13]; Wiener [19].

²⁵ Cf. Borel [4]; Lévy [12], [13]; Wiener [19].

- [5] Borel, E. and Deltheil, R., *Probabilités; erreurs*, 4th edition (Paris), Colin (1934).
- [6] Cramér, H., "Ueber eine Eigenschaft der normalen Verteilungsfunktion," *Mathematische Zeitschrift*, vol. 41 (1936), pp. 405-414.
- [7] Deltheil, R., *Probabilités géométriques* (Paris), Gauthier-Villars (1936).
- [8] Haviland, E. K., "On the inversion formula for Fourier-Stieltjes transforms in more than one dimension," *American Journal of Mathematics*, vol. 57 (1935); I, pp. 94-100; II, pp. 382-388.
- [9] Jessen, B. and Wintner, A., "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.
- [10] Kac, M., "On a characterization of the normal distribution," *American Journal of Mathematics*, vol. 61 (1939), pp. 726-728.
- [11] Lévy, P., "Théorie des erreurs. La loi de Gauss et les lois exceptionnelles." *Bulletin de la Société Mathématique de France*, vol. 52 (1924), pp. 56-58.
- [12] Lévy, P., *Leçons d'analyse fonctionnelle* (Paris), Gauthier-Villars (1922), pp. 262-268, 274-284.
- [13] Lévy, "Analyse fonctionnelle," *Mémoires des Sciences Mathématiques*, fasc. 5 (1925), pp. 39-40.
- [14] Schoenberg, I. J., "On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space," *Annals of Mathematics*, vol. 38 (1937), pp. 787-793.
- [15] Schoenberg, I. J., "Metric spaces and completely monotone functions," *Annals of Mathematics*, vol. 39 (1938), pp. 811-841.
- [16] Schoenberg, I. J., "Metric spaces and positive definite functions," *Transactions of American Mathematical Society*, vol. 44 (1938), pp. 522-536.
- [17] Sommerfeld, A., "Eine besonders anschauliche Ableitung des Gaussischen Fehlergesetzes," *Boltzmann-Festschrift*, Leipzig (1904), pp. 848-859.
- [18] Wiener, N., "Differential space," *Journal of Mathematics and Physics*, Massachusetts Institute of Technology, vol. 2 (1923), pp. 131-174.
- [19] Wiener, N. and Wintner, A., "On singular distributions," *Journal of Mathematics and Physics*, Massachusetts Institute of Technology, vol. 17 (1939), pp. 233-246.
- [20] Wintner, A., "Upon a statistical method in the theory of diophantine approximations," *American Journal of Mathematics*, vol. 55 (1933), pp. 309-331.
- [21] Wintner, A., "On a class of Fourier transforms," *American Journal of Mathematics*, vol. 58 (1936), pp. 45-90 and p. 425.
- [22] Wintner, A., *Spektraltheorie der unendlichen Matrizen*, Leipzig (1929), pp. 81-83, 88-91.
- [23] Wintner, A., "Spherical equidistributions and a statistics of polynomials which occur in the theory of perturbations," *Strömberg-Festschrift*, Copenhagen, 1940 (in press).

ON UPPER LIMIT RELATIONS FOR NUMBER THEORETICAL FUNCTIONS.*

By PHILIP HARTMAN and RICHARD KERSHNER

There are, in the literature, several results on the limit superior of number theoretical (i. e., additive or multiplicative) functions, giving results of the following nature:

$$(1) \quad \limsup_{x \rightarrow \infty} f(x)g(x) = 1,$$

where $f(x)$ is a number theoretical function and $g(x)$ is elementary. All these results have in common the fact that the functions $f(x)$ and $g(x)$ considered are of such a nature that

$$(2) \quad \lim_{n \rightarrow \infty} f(r_n)g(r_n) = 1,$$

where

$$r_n = p_1 p_2 \cdots p_n$$

is the product of the first n primes.

The purpose of this note is to delimit a simple class of functions for which results of this nature can be obtained. This possibility was suggested to us by Professor Wintner. The greater portion of the paper will deal with additive functions; although multiplicative functions may, of course, be treated by applying these results to their logarithms, this consideration leaves something to be desired, since from

$$\log f(x) \leq (1 + \delta)/g(x), \quad x > X(\delta),$$

one can infer only

$$f(x) \leq \exp [(1 + \delta)/g(x)], \quad x > X(\delta),$$

and not

$$f(x) \leq (1 + \delta) \exp (1/g(x)), \quad x > X'(\delta),$$

which would be needed to prove a corresponding limit relation. Correspondingly, the direct treatment of the multiplicative case seems to be more difficult than that of the additive case, and we were unable to establish for multiplicative functions a result of generality comparable to that obtained for additive functions. Thus we have confined the consideration of multiplicative

* Received March 14, 1940.

functions to one very simple case; which, however, does imply the known limit result for the Euler ϕ -function.

The treatment will be based on a very simple lemma stating the Tauberian conditions needed in order to infer (1) from (2).

LEMMA. *Let $f(x)$, $0 < x < +\infty$, be a real-valued function of the integer x , and let $\{r_k\}$ be a sequence of integers, with the following properties:*

- (i) $0 < r_k < r_{k+1}, \quad (k = 1, 2, \dots),$
- (ii) $r_k \rightarrow \infty, \text{ as } k \rightarrow \infty,$
- (iii) $f(r_k)/f(r_{k+1}) \rightarrow 1, \text{ as } k \rightarrow \infty,$
- (iv) *for every $\delta > 0$ there exists an $N = N_\delta$ such that*
- (3) $f(x) \leq (1 + \delta)f(r_n) \text{ whenever } x \leq r_n \text{ and } n > N.$

Let $g(x)$ be a non-increasing function such that

$$(2 \text{ bis}) \quad f(r_n)g(r_n) \rightarrow 1, \text{ as } n \rightarrow \infty.$$

Then

$$\limsup_{x \rightarrow \infty} f(x)g(x) = 1.$$

In order to prove this lemma, notice that the conditions (iii) and (2 bis) imply that

$$(iii \text{ bis}) \quad g(r_{n-1})/g(r_n) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

If x and n are integers such that the relations $r_{n-1} < x \leq r_n$ and (3) hold, then, in virtue of the monotony of the function g ,

$$f(x)g(x) \leq (1 + \delta)[f(r_n)g(r_n)][g(r_{n-1})/g(r_n)].$$

Hence, it follows from (2 bis) and (iii bis) that $\limsup f(x)g(x)$ is not greater than 1. On the other hand, (2 bis) alone implies that it is not less than 1. This completes the proof of the lemma.

Before proceeding to the general class of additive functions mentioned above, to which this lemma is applicable, two special cases which become immediately obvious when thought of in connection with this lemma will be mentioned. These are the cases of strongly additive and strongly multiplicative functions. An additive (multiplicative) function is called strongly additive (strongly multiplicative) if $f(p^\nu) = f(p)$ for all $\nu = 1, 2, \dots$. (Throughout the paper p will denote a prime and p_n the n -th prime.)

THEOREM I. *Let $f(x)$, $x = 1, 2, \dots$, be strongly multiplicative, so that $f(p_k^\nu) = f(p_k)$, $f(p_j p_k) = f(p_j)f(p_k)$, ($j \neq k$). Let $f(p_{k-1}) \geq f(p_k) \rightarrow 1$ as*

$k \rightarrow \infty$. Then the conditions (i)-(iv) of the Lemma are satisfied by the sequence $r_n = p_1 p_2 \cdots p_n$.

The proof is obvious, in fact (3) is satisfied for $N=1$ and $\delta=0$. It should be mentioned that the requirement of monotony, $f(p_{k-1}) \geq f(p_k)$, cannot be dispensed with. This can be seen by the example

$$f(p_{2^n}) = 1 + 1/n,$$

$$f(p_k) = 1 \text{ if } k \neq 2^n \text{ for any } n.$$

In spite of the simplicity of Theorem I, the known case¹ of the function $f(x) = x/\phi(x)$, where $\phi(x)$ is the Euler ϕ -function, may be treated as a particular case of this theorem. In fact, $x/\phi(x)$ is strongly multiplicative and

$$p/\phi(p) = (1 - 1/p)^{-1},$$

so the conditions of Theorem I are satisfied. Consequently, the relation

$$\frac{p_1 p_2 \cdots p_n}{\phi(p_1 p_2 \cdots p_n)} \cdot \frac{1}{e^C \log \log (p_1 p_2 \cdots p_n)} \rightarrow 1, \text{ as } n \rightarrow \infty,$$

(where C is the Euler constant), which is a consequence of Merten's asymptotic formula

$$\prod_{p \leq x} (1 - 1/p)^{-1} \sim e^C \log x$$

and Chebyshev's inequalities, implies by the Lemma, that

$$\limsup_{x \rightarrow \infty} \frac{x}{\phi(x)} \cdot \frac{1}{e^C \log \log x} = 1.$$

The corresponding theorem for the additive case is the following:

THEOREM II. Let $f(x)$, $x = 1, 2, \cdots$, be strongly additive, so that $f(p_k^v) = f(p_k)$, $f(p_j p_k) = f(p_j) + f(p_k)$, ($j \neq k$). Let $f(p_{k-1}) \geq f(p_k) > 0$. Then the conditions (i)-(iv) of the Lemma are satisfied by the sequence $r_n = p_1 p_2 \cdots p_n$.

The proof is again obvious. Notice, in connection with our earlier remarks on the comparative difficulties of the two cases that the requirements of this Theorem II are much weaker than those of the corresponding Theorem I. It might also be mentioned, in this same connection, that in this case the requirement of monotony can be considerably modified.

As an application of Theorem II, consider the strongly additive function $f(n) = f_a(n)$ defined by

¹ E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*, Leipzig (1909), pp. 219-222.

$$f_a(p) = -\log(1 - 1/p^a), \quad f_a(p_{i_1}^{\nu_1} \cdots p_{i_m}^{\nu_m}) = \sum_{k=1}^m f_a(p_{i_k}).$$

The conditions of Theorem II are obviously satisfied. Also, one has

$$f_a(r_n) = \sum_{j=1}^n f_a(p_j) \sim \sum_{j=1}^n 1/p_j^a, \quad (0 < \alpha < 1);$$

but, in virtue of the prime number theorem,²

$$\sum_{p \leq x} 1/p^a \sim \sum_{n=2}^x 1/n^a \log n \sim x^{1-a}(1-\alpha) \log x, \quad (0 < \alpha < 1),$$

so that (2 bis) is satisfied by the function

$$g(n) = (1 - \alpha) \log \log n / (\log n)^{1-a}.$$

Hence, by Theorem II and the Lemma,

$$\limsup_{x \rightarrow \infty} f_a(x) \log \log x / (\log x)^{1-a} = 1/(1 - \alpha), \quad (0 < \alpha < 1).$$

The strongly additive function $f_a(n)$ in this relation may be replaced by the additive function $\log [\sigma_a(n)/n^a]$, where $\sigma_a(n)$ is the classical (multiplicative) function defined as the sum of the α -th power of the divisors of n . For

$$\sigma_a(p^k)/p^{ka} = 1 + 1/p^a + 1/p^{2a} + \cdots + 1/p^{ka},$$

which implies, for $0 < \alpha < 1$, that

$$\log [\sigma_a(p^k)/p^{ka}] < f_a(p^k) = f_a(p) \text{ and } \log [\sigma_a(r_n)/r_n^a] \sim f_a(r_n).$$

Consequently, the result obtained for $f_a(n)$ may be transcribed as

$$\limsup_{x \rightarrow \infty} \log [\sigma_a(x)/x^a] \log \log x / (\log x)^{1-a} = 1/(1 - \alpha), \quad (0 < \alpha < 1),$$

which was first proved by Gronwall³ (using a refinement of the prime number theorem).

As another example of the use of Theorem II, consider the function $f(x) = \rho(x)$ defined as the number of distinct prime divisors of x . It is easily verified that this function is strongly additive. Since $\rho(p_k) = 1$, the condi-

² This is a consequence of the standard procedure of writing

$$\sum_{p \leq x} 1/p^a = \sum_{p \leq X} 1/p^a + \sum_{X < n \leq x} [\theta(n) - \theta(n-1)]/n^a \log n, \text{ where } \theta(x) = \sum_{p \leq x} \log p,$$

applying the Abel summation formula to the last sum, and using the prime number theorem in the form $(1 - \epsilon)n < \theta(n) < (1 + \epsilon)n$, if $n > X$. (Cf., e.g., *loc. cit.*¹, p. 25.)

³ T. H. Gronwall, "Some asymptotic expressions in the theory of numbers," *Transactions of the American Mathematical Society*, vol. 14 (1913), pp. 113-122. Gronwall also considers the functions $\sigma_a(n)/n^a$ for $a \geq 1$. However, these cases are simpler than the ones treated above; in fact, they are easily handled in the multiplicative form, i.e., without resorting to logarithms. On the other hand, the upper limit is not approached on the sequence $r_n = p_1 p_2 \cdots p_n$, as is the situation above.

tions of Theorem II are satisfied by this function $f(x) = \rho(x)$. Consequently, the relation

$$\rho(p_1 p_2 \cdots p_n) \frac{\log \log (p_1 p_2 \cdots p_n)}{\log (p_1 p_2 \cdots p_n)} \rightarrow 1, \quad (n \rightarrow \infty),$$

i. e., the relation

$$(4) \quad \frac{n \log \log (p_1 p_2 \cdots p_n)}{\log (p_1 p_2 \cdots p_n)} \rightarrow 1, \quad (n \rightarrow \infty),$$

which is an easy consequence of the elementary inequalities of Chebyshev, implies, by the Lemma, that

$$(5) \quad \limsup_{x \rightarrow \infty} \rho(x) \frac{\log \log x}{\log x} = 1.$$

We now proceed to the main result.

THEOREM III. *Let $f(x)$, $x = 1, 2, \cdots$, be an additive function such that, for some ϵ , $0 \leq \epsilon < 1$,*

$$(6) \quad f(p_k^v) \leq v^\epsilon, \quad (k = 1, 2, \cdots; v = 2, 3, \cdots);$$

and

$$(7) \quad f(p_k) \rightarrow 1, \quad k \rightarrow \infty.$$

Then the conditions of the Lemma are satisfied with $r_n = p_1 p_2 \cdots p_n$ and $g(x) = \log \log x / \log x$.

Proof. The conditions (i)-(iii) are obviously satisfied. In order to prove (iv), let $\delta > 0$ be fixed and let

$$(8) \quad x = p_{i_1}^{v_1} p_{i_2}^{v_2} \cdots p_{i_k}^{v_k} < r_n = p_1 p_2 \cdots p_n.$$

Now (7) implies that

$$\frac{f(r_n)}{n} = \frac{f(p_1) + \cdots + f(p_n)}{n} \rightarrow 1, \quad n \rightarrow \infty;$$

so that, for any $\eta_1 > 0$, and for sufficiently large n ,

$$(9) \quad (1 + \eta_1)n \geq f(r_n) \geq (1 - \eta_1)n.$$

On the other hand, by (6),

$$f(x) \leq \sum_{m=1}^k v_m^\epsilon,$$

so that

$$f(x) \leq \sum_{m=1}^k (v_m^\epsilon \log^\epsilon p_{i_m}) (\log^{-\epsilon} p_{i_m}).$$

It follows from the inequality of Hölder that

$$(10) \quad f(x) \leq \left(\sum_{m=1}^k \nu_m \log p_{i_m} \right)^\epsilon \left(\sum_{m=1}^k \log^{-\epsilon/(1-\epsilon)} p_{i_m} \right)^{1-\epsilon}.$$

Now, by (8),

$$(11) \quad \sum_{m=1}^k \nu_m \log p_{i_m} \leq \sum_{m=1}^n \log p_m.$$

On the other hand, by the inequality of Chebyshev, for any $\eta_2 > 0$ and for sufficiently large n ,

$$(12) \quad \sum_{m=1}^n \log p_m \leq (1 + \eta_2) n \log n.$$

Also, for any $\eta_3 > 0$ and for sufficiently large n ,

$$(13) \quad \sum_{m=1}^k \log^{-\epsilon/(1-\epsilon)} p_{i_m} \leq \sum_{m=1}^n \log^{-\epsilon/(1-\epsilon)} (m+1) \leq (1 + \eta_3) n \log^{-\epsilon/(1-\epsilon)} n.$$

If (11), (12), and (13) are substituted in (10), one has

$$f(x) \leq (1 + \eta_2)^\epsilon (1 + \eta_3)^{1-\epsilon} (n \log n)^\epsilon (n \log^{-\epsilon/(1-\epsilon)} n)^{1-\epsilon}$$

or

$$(14) \quad f(x) \leq (1 + \eta_2)^\epsilon (1 + \eta_3)^{1-\epsilon} n$$

for n sufficiently large. Combining (9) and (14) gives

$$f(x) \leq (1 - \eta_1)^{-1} (1 + \eta_2)^\epsilon (1 + \eta_3)^{1-\epsilon} f(r_n),$$

where $\eta_1 > 0$, $\eta_2 > 0$, $\eta_3 > 0$ may be chosen arbitrarily small if n is sufficiently large. Thus, for any $\delta > 0$, there is an N_δ such that

$$(15) \quad f(x) \leq (1 + \delta) f(r_n) \text{ if } n > N_\delta.$$

This shows that the condition (iv) of the Lemma is satisfied in the present case.

The fact that the function $g(x)$ in the Lemma may be chosen to be

$$g(x) = \log \log x / \log x$$

follows from (4), in virtue of (9). This completes the proof of Theorem III.

It should be mentioned that, in view of (7), the requirement (6) of Theorem III may be replaced by the condition

$$(6 \text{ bis}) \quad f(p_k^\nu) \leq \nu^\epsilon f(p_k), \quad (\nu, k = 1, 2, \dots),$$

and, in fact, in view of the asymptotic character of the result, (6) or (6 bis) need only be required for sufficiently large k . The same is not true, however, with regard to ν and it is quite easy to construct an example where (6) fails only for $\nu = 2$ but where the result (15) no longer holds if x is chosen of the form $x = p_1^2 p_2^2 \cdots p_n^2$.

It seems that the requirement (6) or (6 bis) is somewhere near the best estimate of its kind which can imply (15). In fact, it is easily seen that if

$$f(p_k^v) > v/(\log v)^{1-\epsilon} f(p_k) \text{ for some } \epsilon > 0,$$

then (15) fails if x is chosen to be a power of 2.

An example which satisfies the conditions of Theorem III is the function $f(x) = \log d(x)/\log 2$, where $d(x) = \sigma_0(x)$ is the number of distinct divisors of x . In this case $f(x)$ is additive and

$$f(p_k^v) = \log(v+1)/\log 2, \quad (v, k = 1, 2, \dots).$$

Thus, the result,

$$\limsup_{n \rightarrow \infty} \log d(x) \log \log x / \log x = \log 2,$$

due to Wigert,⁴ follows from Theorem III and the Lemma.

QUEENS COLLEGE,
UNIVERSITY OF WISCONSIN.

⁴ Cf. *loc. cit.*¹, pp. 219-222.

ON THE PROPERTIES OF A COLLECTIVE.*¹

By Z. W. BIRNBAUM and HERBERT S. ZUCKERMAN.

1. R. v. Mises² gives the following definition of the simplest collective which he also calls an alternative: A simple collective is an infinite sequence of observations, the result of each of which may be represented by one of two symbols, say 0 or 1, which satisfies

Postulate 1. If n_0 and n_1 are the number of observations, among the first n , for which the results are 0 and 1 respectively, then the limits of the relative frequencies, $\lim_{n \rightarrow \infty} n_0/n = w_0$ and $\lim_{n \rightarrow \infty} n_1/n = w_1$ shall exist; and

Postulate 2. If an infinite subsequence of the total sequence is formed by a "selection" then, for this subsequence, the same limits exist and their values remain unchanged, $\lim_{n \rightarrow \infty} n'_0/n = w_0$, $\lim_{n \rightarrow \infty} n'_1/n = w_1$.

The numbers w_0 and w_1 are called probabilities of the appearance of the labels 0 and 1 in the collective.

These postulates have become the object of considerable discussion. Most of these discussions have centred around the second postulate and a number of investigations have been made in attempts to prove the consistency of the concept of a collective, in connection with the difficulties encountered in interpreting this postulate.³

It is the aim of the present paper to prove that a sequence which fulfills the first postulate, fulfills also, generally speaking, the second postulate. The precise formulation of this statement is given in the following

THEOREM A. *The set of all infinite selections can be interpreted as a space \mathfrak{S} in which a Lebesgue measure is defined, so that if a sequence of 0's*

* Received February 21, 1940.

¹ Presented to the American Mathematical Society, February 24, 1940.

² R. v. Mises, *Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und theoretischen Physik*, Leipzig u. Wien 1931, p. 14.

³ Certain special cases of our Theorem A are included in some of these investigations e.g. in A. H. Copeland, "Point set theory applied to the random selection of the digits of an admissible number," *American Journal of Mathematics*, vol. 58 (1936), pp. 181-192. A special case is also formulated by H. Steinhaus, "Les probabilités dénombrables et leur rapport à la théorie de la mesure," *Fundamenta Mathematicae*, vol. 4 (1923), pp. 286-310, especially p. 305.

and 1's fulfills the first postulate of v. Mises, the second postulate is fulfilled for the subsequence determined by every selection with the exception of a set of measure zero in \mathfrak{S} .

Theorem A follows from a more general theorem which will be formulated in the next paragraph.

2. Let K be an infinite sequence (a_1, a_2, \dots) of 0's and 1's. The number of 1's among the first n elements of that sequence is $\sum_{i=1}^n a_i$. To each sequence K we ascribe the real number $k = a_1/2 + a_2/2^2 + \dots$.

Let S be a selection which, if applied to a sequence K , preserves only the i_1 -st, i_2 -nd, \dots terms. The result of an application of S to K is, therefore, the sequence a_{i_1}, a_{i_2}, \dots which we shall denote by $K \subset S$, in accordance with a notation introduced by Copeland.⁴

A selection S is completely described by a sequence (b_1, b_2, \dots) where $b_{i_i} = b_{i_2} = \dots = 1$, and $b_i = 0$ for all other values of i . We obviously have

$$(1) \quad n = \sum_{i=1}^{i_n} b_i.$$

We shall consider only selections S which preserve infinitely many terms of a sequence to which they are applied, i. e. selections S with $b_i = 1$ for infinitely many values of i . A one-to-one correspondence between the set \mathfrak{S} of all such selections S and all real numbers s of the interval $\langle 0, 1 \rangle$ can be established by ascribing to the selection $S = (b_1, b_2, \dots)$ the number $s = b_1/2 + b_2/2^2 + \dots$. We introduce a measure in \mathfrak{S} by calling a set Σ in \mathfrak{S} measurable if and only if the set σ of corresponding numbers in $\langle 0, 1 \rangle$ is measurable in the sense of Lebesgue, and by defining

$$\text{measure of } \Sigma = m(\Sigma) = \text{measure of } \sigma = m(\sigma).$$

The relative frequencies of the 1's in K are

$$(2) \quad f_n(K) = \frac{1}{n} \sum_{i=1}^n a_i$$

while those in $K \subset S$ are given by

$$(3) \quad f_n(K \subset S) = \frac{1}{n} \sum_{i=1}^{i_n} a_i b_i$$

THEOREM B. If $F(K)$ is the set of points of condensation of the sequence $f_1(K), f_2(K), \dots$, then $F(K) = F(K \subset S)$ almost everywhere in \mathfrak{S} , i. e. for all S except those of a set of measure zero.

⁴ loc. cit. ³.

Proof of Theorem B. We denote by $r_i(t)$, $i = 1, 2, \dots$, the well known Rademacher⁵ functions which are defined for $0 \leq t \leq 1$ as follows: if $t = t_1/2 + t_2/2^2 + \dots$ is the infinite dyadic expansion of t , then $r_i(t) = 1$ if $t_i = 1$, and $r_i(t) = -1$ if $t_i = 0$. We evidently have

$$(4) \quad t_i = \frac{1}{2}(r_i(t) + 1).$$

Using as arguments for those functions the numbers k and s which correspond to K and S , we find, from (2) and (4),

$$(5) \quad f_n(K) = \frac{1}{2}\left(1 + \frac{1}{n} \sum_{i=1}^n r_i(k)\right),$$

and, from (1), (3), and (4),

$$(6) \quad f_n(K \subset S) = \frac{\frac{1}{2} \left[\frac{1}{i_n} \sum_{i=1}^{i_n} r_i(k) r_i(s) + \frac{1}{i_n} \sum_{i=1}^{i_n} r_i(s) + \frac{1}{i_n} \sum_{i=1}^{i_n} r_i(k) + 1 \right]}{\frac{1}{i_n} \sum_{i=1}^{i_n} r_i(s) + 1}.$$

The functions $r_1(s), r_2(s), \dots$ are a normed orthogonal system. It is known⁶ that for such a system the relation

$$(7) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N r_i(s) = 0$$

holds for almost all values of s in $\langle 0, 1 \rangle$. Similarly, the functions $\rho_1(s) = r_1(k)r_1(s)$, $\rho_2(s) = r_2(k)r_2(s)$, \dots , form a normed orthogonal system, and therefore we again have

$$(8) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N r_i(k)r_i(s) = 0,$$

for almost all s . From (5), (6), (7), and (8) we see that

$$(9) \quad \lim_{n \rightarrow \infty} \{f_n(K \subset S) - f_{i_n}(K)\} = 0,$$

for almost all S . Hence every point of condensation of the sequence $\{f_n(K \subset S)\}$ is a point of condensation of the sequence $\{f_{i_n}(K)\}$, and therefore $F(K \subset S)$ is contained in $F(K)$ for almost all S .

We shall now prove that $F(K \subset S)$ also contains $F(K)$ for almost all S . It is easy to see that if $F(K)$ contains two numbers $r < u$, then it also contains all numbers t with $r \leq t \leq u$. We let α be the smallest and β the

⁵H. Rademacher, "Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen," *Mathematische Annalen*, vol. 87 (1922), pp. 112-138.

⁶S. Banach, "Sur la valeur moyenne des fonctions orthogonales," *Bull. Ac. Crac.*, 1919, pp. 66-72.

largest number in $F(K)$. It will suffice to prove that α and β belong to $F(K \subset S)$ for almost all S .

Let $\{l_n\}$ be a sequence of such indices that $\lim_{n \rightarrow \infty} f_{l_n}(K) = \alpha$. If α is not contained in $F(K \subset S)$ for a certain S then it is not a point of condensation of $\{f_n(K \subset S)\}$ and, by (9), it is not a point of condensation of $\{f_{i_n}(K)\}$. Therefore only a finite number of the indices i_n are equal to some l_m , and, if S is (b_1, b_2, \dots) , then we have $b_{l_m} = 0$ from a certain m on. We now let T_r be the set of all S such that $b_{l_m} = 0$ for all $m \geq r$, and $T = \sum_{r=1}^{\infty} T_r$. All S for which α is not a point of condensation of $\{f_n(K \subset S)\}$ belong to T . However, it is easily seen that each set T_r is of measure zero, and hence T is also of measure zero. From this we see that the set E_α , of all S for which α is not a point of condensation of $\{f_n(K \subset S)\}$, is of measure zero. By the same argument, the set E_β , of those S for which β is not a point of condensation of the sequence $\{f_n(K \subset S)\}$ is, too, a set of measure zero. If E is the set (of measure zero) of those S for which (9) does not hold, then $E' = \langle 0, 1 \rangle - E - E_\alpha - E_\beta$ is of measure one and contains only selections S for which both α and β belong to $F(K \subset S)$. If both α and β belong to $F(K \subset S)$ then $F(K \subset S)$ contains every number between α and β , and hence contains $F(K)$. Since the measure of E' is one, this completes the proof of Theorem B.

3. Theorem B states that, for a fixed K , there is a set of measure one of selections S which leave the set of points of condensation of the sequence of relative frequencies invariant, i. e. $F(K \subset S) = F(K)$.

The dual statement is also true: ⁷ for a fixed selection S and almost all K we have $F(K) = F(K \subset S)$. To see this we note that by a classical theorem due to Borel,⁸ for almost all K , the set $F(K)$ contains only the number $\frac{1}{2}$. On the other hand from (6) and (7) we find that, for a fixed S and almost all K , we have $\lim_{N \rightarrow \infty} f_n(K \subset S) = \frac{1}{2}$.

The question may be asked whether it is possible to find a set M of sequences K and a set N of selections S such that each set is of measure one and that $F(K) = F(K \subset S)$ for every K in M and every S in N . The answer to this question is negative as may be seen from the following argument:

We first discard the set of measure zero of those K which contain only a

⁷ For a more general treatment of such "dual" problems i. e. those with a fixed selection and sets of collectives, see Z. W. Birnbaum and J. Schreier, "Eine Bemerkung zum starken Gesetz der grossen Zahlen," *Studia Mathematica*, vol. 4 (1933), pp. 85-89.

⁸ E. Borel, "Les probabilités dénombrables et leurs applications arithmétiques," *Rend. Circ. mat. Palermo*, vol. 27 (1909), pp. 247-271.

finite number of 1's. Now, if the same sequence of 0's and 1's is used for K and for S , i. e. if $K = S$, then $K \subset S$ is a sequence consisting only of 1's. Therefore, for every K ,^a we have $f_n(K \subset K) = 1$, $n = 1, 2, \dots$. Hence, if 1 is not a point of condensation of $\{f_n(K)\}$ we have $F(K) \neq F(K \subset K)$. For almost all K the set $F(K)$ contains only the number $\frac{1}{2}$, therefore $F(K) \neq F(K \subset K)$ for almost all K . It follows that, if M is a set of measure one of sequences K , and N a set of selections S such that for all K in M and all S in N we have $F(K) = F(K \subset S)$, then N must not contain any $S = K$ with K contained in M , and therefore the measure of N is zero.

UNIVERSITY OF WASHINGTON,
SEATTLE, WASHINGTON.

ON SYMMETRIC BERNOULLI CONVOLUTIONS.*

By TATSUO KAWATA.

1. Let $\Lambda(t; \sigma)$, $-\infty < t < +\infty$, denote the Fourier-Stieltjes transform

$$(1) \quad \Lambda(t; \sigma) = \int_{-\infty}^{+\infty} e^{itx} d\sigma(x)$$

of a distribution function $\sigma(x)$, $-\infty < x < +\infty$. Let $\beta(x)$ denote the symmetric Bernoulli distribution, which has at either of the points $x = \pm 1$ the jump $\frac{1}{2}$; so that $\Lambda(t; \beta) = \cos t$, and so the Fourier-Stieltjes transform of the distribution function $\beta(x/b)$, where $b > 0$, is $\cos bt$. Thus,¹ the infinite convolution

$$(2) \quad \sigma(x) = \beta(x/b_1) * \beta(x/b_2) * \beta(x/b_3) * \cdots, \text{ where } b_k > 0,$$

is convergent if and only if

$$(3) \quad \sum b_k^2 < \infty,$$

in which case

$$(4) \quad \Lambda(t; \sigma) = \prod_{k=1}^{\infty} \cos b_k t.$$

Wintner has obtained on the one hand² Gaussian estimates of $1 - \sigma(x)$ and $\sigma(-x)$ for large $x > 0$ in case of an arbitrary $\{b_k\}$ satisfying (3), and on the other hand³ almost Gaussian estimates of $\Lambda(t; \sigma) = \Lambda(-t; \sigma)$ for large $t > 0$ in case $\{b_k\}$ is suitably chosen (e. g., $b_k = k^{-\frac{1}{2}+\epsilon}$, $\epsilon > 0$); he has also pointed out² the relation of these estimates to a conjecture of Wiener, proved by Hardy.⁴ The object of this note is a precise investigation of this relation.

2. First, if $\{b_k\}$ satisfies (3), then there exists a $\lambda > 0$ such that²

$$(5) \quad 1 - \sigma(x) = O \exp(-\lambda x^2) \text{ and } \sigma(-x) = O \exp(-\lambda x^2), \text{ as } x \rightarrow +\infty.$$

Actually, (5) holds for every fixed λ . In order to see this, one merely has to combine the proof² for the existence of a sufficiently small λ with a known device,⁵ which consists in replacing the sequence b_1, b_2, \cdots by the sequence b_{N+1}, b_{N+2}, \cdots , where $N = N(\lambda)$.

* Received March 24, 1939.

¹ B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta-function," *Transactions of the American Mathematical Society*, vol. 38 (1935), p. 61.

² A. Wintner, "Gaussian distributions and convergent infinite convolutions," *American Journal of Mathematics*, vol. 57 (1935).

³ A. Wintner, "On analytic convolutions of Bernoulli distributions," *American Journal of Mathematics*, vol. 56 (1934); "On symmetric Bernoulli convolutions," *Bulletin of the American Mathematical Society*, vol. 41 (1935).

⁴ G. H. Hardy, "A theorem concerning Fourier transforms," *Journal of the London Mathematical Society*, vol. 8 (1933).

In the particular case where $\Lambda(t; \sigma)$ is so small for large $|t|$ as to imply the existence of a continuous derivative $\sigma'(x)$, one has

$$(6) \quad \sigma'(x) = O \exp(-\lambda x^2), \quad x \rightarrow \pm \infty,$$

for every fixed λ . This follows from (5) by a known argument.⁵

Now, (6) implies that *there does not exist a convergent Bernoulli convolution (2) whose Fourier-Stieltjes transform (1) is $O \exp(-\delta t^2)$ for a sufficiently small $\delta > 0$.*

In fact, if there existed a $\delta > 0$ for a suitable sequence $\{b_k\}$ satisfying (3), then, on choosing λ in (6) sufficiently large, one could conclude from the theorem of Hardy⁴ that $\Lambda(t; \sigma)$ is of the form $P(t) \exp(-\alpha t^2)$, where $P(t)$ is a polynomial and α a constant. This involves a contradiction, since (4) has infinitely many (real) zeros and does not vanish identically.

3. It will now be shown that the result of Section 2 cannot be essentially improved. In fact, it will be shown that *there exists to every positive increasing function $p(t)$, $0 < t < \infty$, which satisfies the condition*

$$(7) \quad \int_1^\infty \frac{p(t)}{t^5} dt < \infty$$

a convergent symmetric Bernoulli convolution (2) in such a way that

$$(8) \quad \Lambda(t; \sigma) = O \exp(-p(|t|)), \text{ as } t \rightarrow \pm \infty.$$

In the proof it may be assumed that $p(t)$ tends with t to $+\infty$ in a monotonous way, since otherwise we could replace $p(t)$ by $p(t) + t$.

Now put, for $t > 1$,

$$q(t) = \int_1^t \frac{p(u)}{u} du.$$

Then clearly $q(t)$ is increasing and, since

$$\int_t^\infty \frac{p(u)}{u^3} du \geq p(t) \int_t^\infty \frac{du}{u^3} = \frac{p(t)}{2t^2},$$

we have

$$(9) \quad p(t) = o(t^2).$$

Furthermore, since

$$\int_1^t \frac{q(u)}{u^3} du = - \left\{ \frac{q(u)}{2u^2} \right\}_1^t + \frac{1}{2} \int_1^t \frac{p(u)}{u^3} du$$

and

$$\frac{q(t)}{t^2} = \frac{1}{t^2} \int_1^t \frac{p(u)}{u} du = \frac{1}{t^2} \int_1^t o(u) du = o(1),$$

we have

⁵ Cf. B. Jessen and A. Wintner, *loc. cit.*¹, p. 67.

⁶ B. Jessen and A. Wintner, *loc. cit.*¹, p. 68.

$$(10) \quad \int_1^\infty \frac{q(t)}{t^3} dt < \infty.$$

Now

$$(11) \quad q(3t) - q(t) = \int_t^{3t} \frac{p(u)}{u} du \geq p(t) \log 3 \geq p(t) + A$$

for $t \geq t_0$, where $A = \log \frac{1}{\cos(1/3)}$.

Let $r(t)$ denote the inverse function of $q(t)$, and put $\phi(t) = 1/r(t)$. Then we can easily see that $\phi(t) \rightarrow 0$. Since

$$t\phi^2(t) = \frac{t}{r^2(t)} = \frac{q(r(t))}{r^2(t)} = o(1),$$

we have

$$\begin{aligned} \int_1^N \phi^2(t) dt &= N\phi^2(N) - \phi^2(1) - 2 \int_1^N t\phi(t)\phi'(t) dt \\ &= o(1) - \phi^2(1) - 2 \int_1^N t\phi^3(t) \frac{\phi'(t)}{\phi^2(t)} dt \\ &= o(1) - \phi^2(1) + 2 \int_{1/\phi(1)}^{1/\phi(N)} \frac{q(u)}{u^3} du. \end{aligned}$$

Thus,

$$\int_1^\infty \phi^2(u) du < \infty.$$

Since $\phi^2(u)$ is monotone, it follows that (3) is satisfied by $b_n = \phi(nA)$. It will be shown that, for these b_n , the function (4) satisfies (8).

Let $t > 0$. The number of those n which satisfy $b_nt \geq c$ is $\left[\frac{1}{A} q\left(\frac{t}{c}\right) \right]^*$, for the inequality $b_nt \geq c$ is equivalent to $\phi(nA) \geq c/t$, i. e., to $r(nA) \leq t/c$ or $n \leq \frac{1}{A} q\left(\frac{t}{c}\right)$. Thus, the number of those n which satisfy $1 > b_nt \geq 1/3$ is

$$\begin{aligned} [q(3t)/A] - [q(t)/A] &\geq q(3t)/A - q(t)/A - 1 \\ &\geq (p(t) + A)/A - 1 = p(t)/A, \end{aligned}$$

for $t \geq t_0$. Hence,

$$\begin{aligned} |\Lambda(t, \sigma)| &= \left| \prod_{n=1}^\infty \cos(b_nt) \right| \leq \prod_{1 > b_nt \geq 1/3} \cos(b_nt) \\ &\leq (\cos 1/3)^{p(t)/A} = \exp(-p(t)), \text{ for } t \geq t_0. \end{aligned}$$

Since (4) is an even function, the proof of (8) is complete.

Finally, I should like to express my hearty thanks to Professor A. Wintner for his invaluable criticism and advice.

TOHOKU UNIVERSITY,
SENDAI.

* $[x]$ represents the integral part of x .

THE FOUR-VERTEX THEOREM FOR SPHERICAL CURVES.*¹

By S. B. JACKSON.

1. Introduction. The Four-Vortex Theorem or "Vierscheitelsatz" states that every oval of class C'' in the plane possesses at least four extrema of the curvature, where an oval may be defined as a simple closed curve with non-vanishing curvature.² This theorem has been extended to other classes of plane curves by Fog and Graustein³ and to certain restricted classes of space curves by Süss, Takasu, and others.⁴ As regards the space curves, the results have been very fragmentary, and the curves considered have been principally those that are closely enough related to plane curves so that analogous proofs can be carried over. This is not surprising when one considers that the property of closure for a space curve puts a much lighter restriction on the curvature than does the same condition for a plane curve, which is completely determined by the curvature as a function of the arc. Accordingly, it seems more reasonable to look for a generalization of the theorem to spherical curves, with curvature replaced by geodesic curvature, since a curve on the sphere is completely determined by its geodesic curvature as a function of the arc length. Such a generalization is the object of the present paper.

By a suitably chosen inversion, any spherical curve can be transformed into a plane curve. Under this transformation, it is found (§ 3) that the geodesic vertices of the spherical curve, that is, the extrema of the geodesic curvature, are transformed into the vertices of the plane curve. From the known results for plane curves³ there follows at once the existence of at least four geodesic vertices on any simple closed spherical curve of class C''' .

* Received February 19, 1940.

¹ Presented to the Society, April 8, 1938.

² First published apparently by Mukhopadhyaya, "New methods in the geometry of a plane arc," *Bulletin of the Calcutta Mathematical Society*, vol. 1 (1909), pp. 31-37, and since then appearing repeatedly in the literature.

³ D. Fog, "Über den Vierscheitelsatz und seine Verallgemeinerungen," *Sitzungsberichte der Berlin Akademie der Wissenschaft* (1933), pp. 251-254; W. C. Graustein, "Extensions of the four-vertex theorem," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 9-23.

⁴ W. Süss, "Ein Vierscheitelsatz bei geschlossenen Raumkurven," *Tôhoku Mathematical Journal*, vol. 29 (1928), pp. 359-362; T. Takasu, "Vierscheitelsatz für Raumkurven," *Tôhoku Mathematical Journal*, vol. 39 (1934), pp. 292-298. Also a number of other papers. W. C. Graustein and S. B. Jackson, "The four-vertex theorem for a certain type of space curves," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 737-741.

In pushing the results beyond the case of the simple closed curves, a study of certain spherical arcs is made, called arcs of type Ω (§ 5) because of their shape. These are entirely analogous to Graustein's arcs of type Ω in the plane. It turns out, in fact, that by a suitable inversion a spherical arc of type Ω may be carried into a plane arc of type Ω . Thereby the fundamental property of the plane arcs of type Ω is transferred at once to the spherical arcs of type Ω . This property states that there exists at least one non-negative minimum of geodesic curvature interior to any spherical arc of type Ω . By means of it the Four-Vertex Theorem is extended to a large class of non-simple spherical curves.

In a paper in 1936,⁵ Graustein strengthened the original Four-Vertex Theorem. A vertex is called primary if, at the vertex, the curvature is greater than or less than the average curvature according as it is a maximum or a minimum, and it is shown that the primary vertices outnumber the other (secondary) vertices by at least four for every plane oval. We shall establish precisely analogous results for a certain class of spherical curves, namely those which are tangent indicatrices of other spherical curves (§ 7). The question as to whether the strengthened theorem holds for a wider class of spherical curves is left open.

A close relationship is exhibited between the geodesic vertices on the tangent indicatrix of a twisted space curve, and the dual vertices defined by Takasu⁴ (§ 8). The relation of the geodesic vertices of a spherical curve to the ordinary vertices, that is, the extrema of the ordinary curvature, is also clarified (§ 9). It appears that every geodesic vertex is a vertex, but not conversely, whence any spherical curve has at least as many vertices as geodesic vertices.

2. Transformation of curves by inversion. If $C: x = x(s)$ is a regular twisted space curve of class C''' , lying on a surface, Σ , the following well known equations are valid:⁶

$$(2.1) \quad \begin{aligned} \frac{d\alpha}{ds} &= \frac{v}{\rho} + \frac{\xi}{\tau} \\ \frac{dv}{ds} &= -\frac{\alpha}{\rho} - \frac{\xi}{\tau} \\ \frac{d\xi}{ds} &= -\frac{\alpha}{r} + \frac{v}{\tau} \end{aligned}$$

⁵ W. C. Graustein, "A new form of the four-vertex theorem," *Monatshefte für Mathematik und Physik*, Wirtinger Festband (1936), pp. 381-384.

⁶ See, for example, W. C. Graustein, *Differential Geometry*, Macmillan (1935), pp. 163-165.

where $1/\rho$, $1/r$, and $1/\tau$ are, respectively, the geodesic curvature, the normal curvature, and the geodesic torsion of C on Σ , and α , ξ , v are, respectively, the unit tangent vector to C , the unit normal vector to Σ , and the unit vector tangent to Σ and orthogonal to C such that $(\alpha v \xi) = 1$.⁷

Let us seek the equations of transformation for the quantities $1/\rho$, $1/r$, $1/\tau$ and the curvature $1/R$, of C under an inversion in space. If the sphere of inversion has radius a and center O , the equation of the inversion in vector form is

$$(2.2) \quad x' = \frac{a^2 x}{(x|x)} \quad \text{or, inversely,} \quad x = \frac{a^2 x'}{(x'|x')},$$

where x and x' denote the vectors \overrightarrow{OP} and $\overrightarrow{OP'}$, respectively. From this equation it follows that the relation between the elements of arc, ds and ds' , of C and its image C' , respectively, is

$$(2.3) \quad \frac{ds}{ds'} = \frac{(x|x)}{a^2} = \frac{a^2}{(x'|x')}.$$

If δ is an arbitrary unit vector localized at the point, P , and δ' is the corresponding unit vector at the inverse point, P' , it is readily shown that

$$(2.4) \quad \delta' = \delta - \frac{2(x|\delta)}{(x|x)} x.$$

In particular, the vectors α , v , ξ of the trihedral of C on Σ transform into

$$(2.5) \quad \begin{aligned} \alpha' &= \alpha - \frac{2(x|\alpha)}{(x|x)} x \\ v' &= v - \frac{2(x|v)}{(x|x)} x \\ \xi' &= \xi - \frac{2(x|\xi)}{(x|x)} x. \end{aligned}$$

The vectors α' and v' may be viewed as the first two vectors of the trihedral of the inverted curve, C' , on the inverted surface, Σ' . Since inversion carries a right trihedral into a left trihedral and vice versa, it is necessary to take for the surface normal to Σ' not ξ' but $\xi'' = -\xi'$ in order to preserve the convention that the trihedral have the same disposition as the axes. The trihedral for C' on Σ' is, therefore, α' , v' , ξ'' and equations (2.1) for C' become

$$(2.6) \quad \begin{aligned} \frac{d\alpha'}{ds'} &= \frac{v'}{\rho'} + \frac{\xi''}{\tau'} \\ \frac{dv'}{ds'} &= -\frac{\alpha'}{\rho'} - \frac{\xi''}{\tau'} \\ \frac{d\xi''}{ds'} &= -\frac{\alpha'}{\tau'} + \frac{v'}{\rho'} \end{aligned}$$

where the primes denote quantities referred to C' .

⁷ For vector notation see Chapter I, *loc. cit.* 6.

Differentiating the first of (2.5) with respect to s' and substituting from (2.1), (2.3), and (2.6), we obtain the relation

$$(2.7) \quad \frac{v'}{\rho'} + \frac{\xi''}{r'} = \frac{(x|x)}{a^2} \frac{1}{\rho} v + \frac{(x|x)}{a^2} \frac{1}{r} \xi - \frac{2(x|\alpha)}{a^2} \alpha - \frac{2(x|v)}{a^2} \frac{1}{\rho} x \\ - \frac{2(x|\xi)}{a^2} \frac{1}{r} x - \frac{2}{a^2} x + \frac{4(x|\alpha)^2}{a^2(x|x)} x.$$

The inner product of (2.7) with the second of formulas (2.5) yields the equation

$$(2.8) \quad \frac{1}{\rho'} = \frac{(x|x)}{a^2} \frac{1}{\rho} + \frac{2(x|v)}{a^2}$$

which represents the desired transformation of the geodesic curvature of C into the geodesic curvature of C' . By differentiation of (2.8) with respect to s' and substitution from (2.1) and (2.3), we find

$$(2.9) \quad \frac{d}{ds'} \left(\frac{1}{\rho'} \right) = \left(\frac{ds}{ds'} \right)^2 \frac{d}{ds} \left(\frac{1}{\rho} \right) - \frac{2(x|\xi)}{a^2} \frac{ds}{ds'} \frac{1}{\tau}$$

as the equation of transformation of the derivative of the geodesic curvature.

In order to obtain the corresponding formulas of transformation for the normal curvature, it is only necessary to form the inner product of (2.7) with $\xi'' = -\xi'$. The result is

$$(2.10) \quad \frac{1}{r'} = -\frac{(x|x)}{a^2} \frac{1}{r} - \frac{2(x|\rho)}{a^2}$$

and differentiation of this relation and use of (2.1) and (2.3) yield the equation

$$(2.11) \quad \frac{d}{ds'} \left(\frac{1}{r'} \right) = -\left(\frac{ds}{ds'} \right)^2 \frac{d}{ds} \left(\frac{1}{r} \right) - \frac{2(x|v)}{a^2} \frac{ds}{ds'} \frac{1}{\tau}.$$

A similar procedure in the case of the geodesic torsion gives the following equation of transformation

$$(2.12) \quad \frac{1}{\tau'} = -\frac{ds}{ds'} \frac{1}{\tau}.$$

Since $1/R^2 = 1/\rho^2 + 1/r^2$, we have, on squaring and adding (2.8) and (2.10)

$$\frac{1}{R'^2} = \left(\frac{ds}{ds'} \right)^2 \frac{1}{R^2} + \frac{4}{a^2} \frac{ds}{ds'} \left(x \left| \frac{v}{\rho} + \frac{\xi}{r} \right. \right) + \frac{4[(x|v)^2 + (x|\xi)^2]}{a^4}.$$

From (2.1) and the Frenet-Serret formulas, it follows that

$$\frac{\beta}{R} = \frac{d\alpha}{ds} = \frac{v}{\rho} + \frac{\xi}{r}$$

where β is the unit principal normal vector for C . Making use of this, together with the identity

$$(x|x) = (x|\alpha)^2 + (x|v)^2 + (x|\xi)^2$$

we find

$$(2.13) \quad \frac{1}{R'^2} = \left(\frac{ds}{ds'}\right)^2 \frac{1}{R^2} + \frac{4}{a^2} \frac{ds}{ds'} \left[1 + \frac{(x|\beta)}{R}\right] - \frac{4(x|\alpha)^2}{a^4}$$

as the equation of transformation for the curvature.

Differentiation of (2.13) and application of the Frenet-Serret formulas gives the equation of transformation for the derivative of $1/R$, namely:

$$(2.14) \quad \frac{1}{R'} \frac{d}{ds'} \left(\frac{1}{R'}\right) = \left(\frac{ds}{ds'}\right)^2 \left\{ \frac{d}{ds} \left(\frac{1}{R}\right) \left[\frac{1}{R} \frac{ds}{ds'} + \frac{2(x|\beta)}{a^2}\right] - \frac{2(x|\gamma)}{a^2} \frac{1}{RT} \right\}$$

where γ is the unit binormal vector for C .

It is to be observed that equations (2.13) and (2.14) are independent of the surface, Σ , since they involve only intrinsic properties of the curve.

As a consequence of the formulas developed above, the following theorem may be stated at once.

THEOREM 2.1. *If a surface, Σ , is carried by inversion into a surface, Σ' .*

(a) *the extrema of geodesic curvature, not at the center of inversion, on the lines of curvature of class C''' of Σ are carried into the similar extrema of geodesic curvature on the corresponding lines of curvature of Σ' , points of maximum (minimum) geodesic curvature going into points of maximum (minimum) geodesic curvature;*⁸

(b) *the extrema of normal curvature, not at the center of inversion, on the lines of curvature of class C''' of Σ are carried into the similar extrema of normal curvature on the corresponding lines of curvature of Σ' , points of maximum (minimum) normal curvature going into points of minimum (maximum) normal curvature.*⁸

The proof is immediate, for the lines of curvature are characterized by the fact that $1/\tau = 0$. Since $ds/ds' \neq 0$, it follows by (2.9) that $d(1/\rho)/ds$ and $d(1/\rho')/ds'$ pass through zero together and in the same direction. This proves (a), since the extrema of geodesic curvature are characterized by the fact that at these points (or arcs) the derivative of the geodesic curvature changes sign, and the direction of passing through zero for the derivative

⁸ These statements as to exactly what the maximum and minimum points are transformed into are valid only by virtue of our agreement regarding the relative orientations of Σ and Σ' .

determines whether it is a maximum or a minimum point.⁹ By use of (2.11), the proof of (b) follows in a similar manner, except that in this case $d(1/r)/ds$ and $d(1/r')/ds'$ pass through zero in opposite directions, so that maximum points are carried into minimum points and vice versa.

3. Geodesic vertices on spherical curves. An extremum of geodesic curvature will be called a geodesic vertex, and a point (or arc) where the geodesic curvature changes sign a geodesic inflection. The term vertex, alone, will be used to indicate an extremum of the curvature, $1/R$. It is necessary to clarify this ambiguous term, curvature, however. For a twisted space curve, C , the curvature is defined as inherently non-negative. For a plane curve, however, we shall use the same word, curvature, to denote what is actually the geodesic curvature of the curve with respect to the plane. This curvature may be either positive or negative depending on the direction of rotation of the tangent with reference to the orientation of the plane.

According to (2.9), geodesic vertices of a curve, C , are preserved under inversion provided $1/\tau = 0$, as was seen in the proof of Theorem 2.1. Special interest thus attaches to those surfaces for which $1/\tau \equiv 0$, i. e., for which all curves are lines of curvature. It is well known that the only such surfaces are the sphere and the plane. Henceforth we shall limit most of our attention to such curves. Part (b) of Theorem 2.1 becomes trivial for such curves, but part (a) assumes the following form.

THEOREM 3.1. *The geodesic vertices, not at the center of inversion, on a plane or spherical curve of class C''' are carried by inversion into the similar geodesic vertices of the transformed curve, points of maximum (minimum) geodesic curvature being carried into points of maximum (minimum) geodesic curvature.⁸*

A simple, closed spherical curve, C , of class C''' , may be carried by a suitably chosen inversion into a simple closed plane curve, \bar{C} , of class C''' . Since every simple closed plane curve of class C''' , not a circle, has at least four vertices³ we obtain at once the following theorem.

THEOREM 3.2. *A simple closed spherical curve, of class C''' , not a circle, has at least four geodesic vertices.¹⁰*

⁹ This proof holds only for isolated extrema, since the first derivative test may fail for extrema which are limit points of other extrema. The theorem is valid for this type of extrema also, but the proof is omitted as of scant interest for the present paper.

¹⁰ This result is incorrectly stated by Fog, *loc. cit.* 3 in that he states that vertices and geodesic vertices coincide on a spherical curve. This is incorrect. (See § 9).

4. *D*-arcs and *D*-curves. It is necessary to introduce at this point a series of lemmas dealing with certain types of spherical arcs and curves. The results and methods parallel very closely certain work by Fenchel on spherical arcs.¹¹

A simple spherical arc of class C' will be called a *D*-arc if (a) it consists of a finite succession of arcs of class C''' with geodesic curvatures continuous clear to their endpoints, and (b) the geodesic curvature is non-negative when the arc is suitably directed. A simple closed spherical curve is called a *D*-curve if every sub-arc of it is a *D*-arc. Otherwise expressed, a *D*-curve is a *D*-arc that is closed. It follows at once from these definitions that on a *D*-arc or *D*-curve the geodesic curvature is continuous except for a finite number of points where one-sided limits exist. In the work that follows we shall consider the sphere as oriented by viewing it from the tip of the outward drawn normal.

LEMMA 4.1. *In a sufficiently small neighborhood of any point on a *D*-arc, the arc lies on or to the left of the directed tangent great circle at this point.*

At a point of continuity of $1/\rho$ the lemma follows from the definition of non-negative geodesic curvature, while at a point of discontinuity of $1/\rho$ the lemma holds for each of the two arcs class C''' which meet at this point and thus holds here also.

LEMMA 4.2. *If a *D*-arc joins two non-diametral points, *A* and *B*, of a great circle and does not meet it elsewhere, the region (contained in a hemisphere) bounded by the *D*-arc and the smaller great circle segment, *AB*, lies to the left of the *D*-arc.*

The proof given by Fenchel¹¹ for an arc of continuous non-vanishing geodesic curvature holds without alteration in the present case. It may be noted, however, that we have assumed *A* and *B* non-diametral, whereas Fenchel could prove it.

LEMMA 4.3. *A *D*-arc, contained in a hemisphere, joining two diametral points, *A* and *B*, is a great semicircle.*

Consider the great semicircle *APB* directed from *A* to *B*, where *P* is any point of the *D*-arc. In case *P* coincides with *A*(*B*) we shall mean by *APB* the semicircle from *A* to *B* which is tangent to the *D*-arc at *A*(*B*). For some point (or points) *P* the *D*-arc lies entirely on or to the right of this great

¹¹ W. Fenchel, "Über Krümmung und Windung geschlossene Raumkurven," *Mathematische Annalen*, vol. 10 (1929), pp. 238-252.

semicircle. The points common to the D -arc and this great semicircle APB are a closed set, and consist entirely of points of tangency, except perhaps for A and B . Moreover, in each case, the tangency must be in the direction APB since otherwise the D -arc must either cross APB or cut itself, both of which are impossible. If the lemma is false there exists at least one such point of tangency in every neighborhood of which there are points of the D -arc to the right of APB . But this contradicts Lemma 4.1 and is therefore impossible.

LEMMA 4.4. *A D -arc has at most a finite number of crossings with any great circle.*

By a crossing is meant any point or arc common to the D -arc and the great circle in every neighborhood of which lie points on both sides of the great circle. Fenchell¹¹ has proved this lemma for arcs of continuous, non-vanishing, geodesic curvature, and this proof extends at once to D -arcs. It should be observed that by Lemma 4.1 a tangency cannot be a crossing.

The remaining essential properties are most readily obtained by considering first the case of D -arcs with non-vanishing geodesic curvature. At a point of discontinuity of $1/\rho$ we demand also that both the one-sided limits shall be different from zero.

LEMMA 4.5. *The tangent great circle to a D -arc of non-vanishing geodesic curvature at a point P has no further points of contact with the D -arc in a sufficiently small neighborhood of P .*

In general $1/\rho = d\phi/ds$ where $\Delta\phi$ is, to within infinitesimals of higher order, the angle between two neighboring tangent great circles. At a point at which $1/\rho$ is continuous, $d\phi > 0$ and the arc is actually turning away from the tangent, while a point at which $1/\rho$ is discontinuous is the junction of two arcs, each of which is turning away from the tangent.

LEMMA 4.6. *The tangent great circle to a D -curve with non-vanishing geodesic curvature, at a point P , has no further points in common with the curve.*

The number of points common to the circle and the curve is finite, for by Lemma 4.4 the number of crossings is finite, and by Lemma 4.5 the closed set of tangencies consists only of isolated points and is therefore finite. Assume that there are common points, other than P , and let Q be the last such point before P . Since $1/\rho > 0$, it follows from Lemma 4.3 that P and Q are not diametral. The tangency at P determines a directed great circle arc, PQ .

Let R denote the region, contained in a hemisphere, bounded by the D -arc QP and the great circle arc PQ . R and the arc QP are on the same side of the great circle arc PQ , and since, by Lemma 4.5, QP lies to the left of PQ at P , R lies to the left of PQ . Since PQ and QP are similarly directed at P , R lies also to the left of the D -arc QP . It follows readily from Lemma 4.2 that PQ is the shorter great circle arc joining P and Q . At P the D -curve actually passes into the interior of R , by Lemma 4.5, and therefore, in order to return to Q , it must return to a point of the great circle arc PQ . This is impossible, by Lemma 4.2 since the region bounded would be on the right. Thus we obtain a contradiction and the lemma is proved.

Fenchell¹¹ obtained the following result, restated here only for convenience.

LEMMA 4.7. *If A is an arc of class C''' on a surface of positive Gaussian Curvature, a similarly directed geodesic parallel to A contained in the field of geodesics perpendicular to A and lying to the left of A has greater (algebraic) geodesic curvature than A at corresponding points.*

Since by this lemma a geodesic parallel to a D -curve which lies sufficiently near it and to its left is surely a D -curve we are led at once to:

LEMMA 4.8. *The geodesic parallels to a D -curve that lie sufficiently near it and to its left are D -curves of non-vanishing geodesic curvature.*

LEMMA 4.9. *A tangent great circle to a D -curve cannot cross the curve.*

Since, by Lemma 4.1, a tangency cannot be a crossing, the curve and a great circle meet at some angle, not zero, at a crossing. Suppose there exists a tangent great circle that crosses the curve. If the D -curve is deformed to its left into an arbitrarily near geodesic parallel, the crossing points and the tangent great circle deform continuously, and the geodesic parallel has a crossing with a tangent great circle. Since this contradicts Lemma 4.6, the assumption is false and the lemma is proved.

It is clear from this lemma that every D -curve is contained in a closed hemisphere, which leads to the following result.

LEMMA 4.10. *A D -curve containing two diametral points is a great circle.*

Since the entire curve, and hence each of the arcs into which the diametral points divide it, is contained in a hemisphere, it follows by Lemma 4.3 that each arc is a great semicircle. The conclusion then follows by the continuity of the tangent.

It can be shown by further discussion that the D -curves have the character of ovals on the sphere. In particular, a D -curve, not a great circle, has in common with any tangent great circle either a single point or a single arc, less than a semicircle. Since this property is not essential for our work, the details of the discussion will be omitted.

5. Arcs of type Ω . Graustein³ has developed a theory of certain plane arcs which he has called arcs of type Ω . We shall consider analogous spherical arcs which will also be designated as of type Ω .

A spherical arc, AB , of class C''' , is said to be of type Ω if (a) its geodesic curvature, when it is traced from A to B , is non-negative and is not identically zero; (b) the tangent great circles at A and B coincide; (c) the arc meets this common tangent only at A and B ; and (d) it is simple except that B may coincide with A . Condition (c) is not actually necessary as a part of the definition since it essentially follows from the other three conditions and the work of § 4. However, it is convenient and we shall retain it.

An arc of type Ω , which may be designated without ambiguity by Ω , is clearly a D -arc. Moreover, it is tangent to the common tangent great circle in the same direction at A and B , since otherwise Lemma 4.1 would be violated at one point or the other. By adjoining to Ω the great circle arc BA , directed in the sense induced by Ω , there arises a D -curve, $\bar{\Omega}$, with discontinuities in the geodesic curvature at A and B . Since Ω is not a great circle arc, $\bar{\Omega}$ is not a great circle, and by Lemma 4.10 the arc BA is less than a semicircle. From this discussion and Lemma 4.9 we have at once the following result.

LEMMA 5.1. *The closure, $\bar{\Omega}$, of an arc of type Ω lies on one side of every tangent great circle. The common tangent great circle at A and B has just the contact arc BA , less than a semicircle, in common with $\bar{\Omega}$.*

Consider the point M' diametrically opposite to a point M on the great circle arc, BA , of $\bar{\Omega}$. Let M' , which by Lemma 5.1 does not lie on $\bar{\Omega}$, be chosen as center of stereographic projection. The great circle containing the arc BA goes into a straight line, and Ω goes into an arc Ω' , lying on one side of this line and tangent to it at the projected points A' and B' . Consider the great circle K tangent to Ω at any point other than A or B . The point M' lies on one side of K , while M , and with it all of $\bar{\Omega}$, lies on the other side by Lemma 5.1, since the circle K by hypothesis is not the common tangent great circle at A and B . Thus K projects into a circle K' with Ω' in its interior. Since, at the point of tangency, Ω' must have at least as great curvature as K' , Ω' has

non-vanishing curvature, except perhaps at A' and B' . It is seen at once that Ω' is a plane arc of type Ω as defined by Graustein.³

The direction on a plane (spherical) arc of type Ω for which the (geodesic) curvature is non-negative will be called the positive direction. It is readily shown that the positive directions on Ω and Ω' correspond. When Ω is traced so that $1/\rho \geq 0$, v is directed toward the interior of $\bar{\Omega}$, that is, toward the smaller of the two simply connected regions into which $\bar{\Omega}$ divides the sphere. Since M' is not in this interior, the interior of $\bar{\Omega}$ projects into the interior of $\bar{\Omega}'$, and v projects into the vector v' directed toward the interior of $\bar{\Omega}'$. But for Ω' the first of formulas (2.6) becomes $d\alpha'/ds' = v'/R'$. This shows that $1/R' \geq 0$ when v' is directed toward the interior of $\bar{\Omega}'$, and the positive directions on Ω and Ω' correspond. We have therefore established

LEMMA 5.2. *By a suitable inversion, a spherical arc of type Ω can be transformed into a plane arc of type Ω , so that the positive directions on the two arcs correspond.*¹²

From Lemma 5.2, Theorem 3.1, and Graustein's theorem³ that the non-negative curvature of a plane arc of type Ω has at least one minimum interior to the arc or is constant throughout the arc, we have at once the following theorem.

THEOREM 5.1. *A spherical arc of type Ω has a minimum of non-negative geodesic curvature interior to the arc, or has constant geodesic curvature throughout the arc.*

This leads readily to a second result, since an arc of type Ω with constant geodesic curvature is a circle.

THEOREM 5.2. *A closed spherical curve of class C''' which has geodesic inflections and contains an arc of type Ω , not a circle, has at least four geodesic vertices.*

If the curve is directed so that on the arc of type Ω , $1/\rho \geq 0$, it follows from Theorem 5.1 that there exists at least one non-negative minimum of $1/\rho$. Since there are geodesic inflections, $1/\rho$ becomes negative, and there must also

¹² This lemma is established only by virtue of the relative orientations of Σ and Σ' agreed on when we derived formulas (2.6). In the present case it implies that if the sphere is oriented by the outward drawn normals, the plane is oriented by the normals directed away from the sphere.

exist a negative minimum. It follows that there must also be two maxima, and thus at least four vertices.

6. Extrema of the torsion of spherical curves. The torsion of a curve C on a surface Σ is readily found to be

$$(6.1) \quad \frac{1}{T} = \frac{\frac{1}{r} \frac{d}{ds} \left(\frac{1}{\rho} \right) - \frac{1}{\rho} \frac{d}{ds} \left(\frac{1}{r} \right)}{\frac{1}{\rho^2} + \frac{1}{r^2}} + \frac{1}{\tau}$$

by replacing the angle which enters into Bonnet's formula, $1/T = d\phi/ds + 1/\tau$, by its value in terms of $1/r$ and $1/\rho$. For a spherical curve on a sphere of radius b , $1/r = -1/b$, $d(1/r)/ds = 0$, $1/\tau = 0$, and (6.1) reduces to

$$(6.2) \quad \frac{1}{T} = \frac{-\frac{1}{b} \frac{d}{ds} \left(\frac{1}{\rho} \right)}{\frac{1}{\rho^2} + \frac{1}{b^2}}.$$

Since the geodesic vertices are the points where $d(1/\rho)/ds$ changes sign, we obtain the following result directly from (6.2).

LEMMA 6.1. *On a spherical curve of class C''' the geodesic vertices are precisely the points where the torsion changes sign.*

Let us call a point (or arc) where $1/T$ changes sign a transition of the torsion. It has been proved by Fenchel for a closed spherical curve of class C''' that $\int_C ds/T = 0$.¹³ Hence, for a closed spherical curve, a transition of the torsion is, in reality, a point at which the torsion crosses its average value. A maximum (minimum) point of the torsion on a closed spherical curve where $1/T > 0$ (< 0) will be called a primary extremum. Other extrema will be termed secondary.¹⁴

LEMMA 6.2. *If C is any closed spherical curve of class C''' , $p_T = s_T + g$, where p_T and s_T are, respectively, the numbers of primary and secondary ex-*

¹³ W. Fenchel, "Über einen Jacobischen Satz der Kurventheorie," *Tôhoku Mathematical Journal*, vol. 39 (1934), pp. 95-97. This result also follows by integrating (6.2).

¹⁴ Compare W. C. Graustein, *loc. cit.* 5. Also W. C. Graustein and S. B. Jackson, *loc. cit.* 4.

trema of the torsion, and g is the number of geodesic vertices on C . It is understood that if s_T or g is infinite the equation merely implies that p_T is also infinite.¹⁴

By Lemma 6.1, g represents the number of transitions of $1/T$ on C . If $1/T \equiv 0$, all three quantities are zero and the formula is trivially valid. In the contrary case there exist at least two transitions. Between two consecutive transitions, the primary extrema outnumber the secondary by one or both are infinite, for on this arc the types of extrema alternate, and the first and last are primary. Finally if g is infinite then p_T is also infinite, since between two transitions there is at least one primary extremum. Thus the lemma holds in every case.

The following theorem is a direct consequence of Lemma 6.2 and Theorem 3.2.

THEOREM 6.1. *The number of primary extrema of the torsion on a simple closed spherical curve of class C''' , not a circle, exceeds the number of secondary extrema by at least four, or both are infinite.*¹⁵

7. Geodesic vertices on tangent indicatrices of spherical curves. If C is a closed spherical curve of length l , a maximum (minimum) point of $1/\rho$ at which $1/\rho - 1/a > 0$ (< 0) will be called a primary geodesic vertex, where $1/a = (1/l) \int_C ds/\rho$; i. e. $1/a$ is the average value of $1/\rho$ taken over C . All other geodesic vertices will be termed secondary. A point (or arc) where $1/\rho - 1/a$ changes sign will be called a transition of $1/\rho$. Precisely as in Lemma 6.2, it can be shown that

$$(7.1) \quad p = s + t$$

where p , s , and t are respectively, the numbers of primary and secondary geodesic vertices, and the number of transitions of $1/\rho$ on C .

It has been shown by Fenchel¹³ that for a regular space curve C_0

$$(7.2) \quad \frac{1}{T_0} = \frac{1}{R_0} \frac{1}{\rho}$$

¹⁵ The existence of at least four transitions of the torsion for any simple closed curve on an ovaloid was proved by H. Mohrmann, "Die Minimalzahl der stationären Ebenen eines räumlichen Ovals," *Sitz. der Königlich Bayerischen Akad. der Wissenschaften, Math.-Phys. Klasse*, München (1917), pp. 51-53. This might have been used in conjunction with Lemma 6.1 to establish Theorem 3.2.

where $1/\rho$ is the geodesic curvature of the tangent indicatrix, or, equivalently, $ds_0/T_0 = ds/\rho$, where s_0 and s are the arc lengths on C_0 and its tangent indicatrix, C , respectively. If, in particular, C_0 is itself a closed spherical curve, $\int_C ds/\rho = \int_C ds_0/T_0 = 0$, as was noted in § 6, whence it follows that a necessary condition for a closed spherical curve, C , to be the tangent indicatrix of some other closed spherical curve, C_0 , is that $\int_C ds/\rho = 0$; i. e. $1/a = 0$.

Since, for a spherical curve C_0 , $1/R_0 \neq 0$, it follows that in this case $1/\rho$ and $1/T_0$ have corresponding transitions, and t , in (7.1), equals the number of transitions of $1/T_0$ which, by Lemma 6.2, equals the number of geodesic vertices on C_0 . Thus we have proved the following theorem.

THEOREM 7.1. *If C_0 is a closed spherical curve of class C''' , and C is its tangent indicatrix, then*

$$p = s + g_0$$

where p and s are the numbers of primary and secondary geodesic vertices on C , respectively, and g_0 is the number of geodesic vertices on C_0 .

Let C_0 be a closed spherical curve of class C^{n+2} and consider the sequence of closed spherical curves C_i , $i = 1, \dots, n$, such that C_i is the tangent indicatrix of C_{i-1} . C_i will be called the i -th tangent indicatrix of C_0 . It may be noted that C_1 and C_2 are the ordinary tangent and normal indicatrices of C_0 . The last theorem may now be generalized in the following way.

THEOREM 7.2. *If C_0 is a closed spherical curve of class C^{n+2} and C_i is its i -th tangent indicatrix, $i = 1, \dots, n$, then*

$$p_r = s_r + 2 \sum_{i=1}^{r-1} s_i + g_0 \quad r \leq n$$

where p_i and s_i are, respectively, the numbers of primary and secondary geodesic vertices on C_i , and g_0 is the number of geodesic vertices on C_0 .

The proof is by induction on r . For $r = 1$ the above contention is precisely Theorem 7.1. If the theorem is true for $r = t$, then $p_t = s_t + 2 \sum_{i=1}^{t-1} s_i + g_0$, and the number of geodesic vertices on C_t is $p_t + s_t = 2 \sum_{i=1}^t s_i + g_0$. Since C_{t+1} is the tangent indicatrix of C_t , it follows by Theorem 7.1 that $p_{t+1} = s_{t+1} + 2 \sum_{i=1}^t s_i + g_0$, and the induction is complete.

COROLLARY 7.2.1. *The k -th tangent indicatrix of a closed spherical curve, C_0 , contains at least as many geodesic vertices as C_0 . The numbers are equal if and only if the first k tangent indicatrices have only primary vertices.*

The proof is immediate, for $p_k + s_k = 2 \sum_{i=1}^k s_i + g_0 \geq g_0$, and the equality sign holds if and only if $s_i = 0$, $i = 1, \dots, k$.

COROLLARY 7.2.2. *The number of primary geodesic vertices on a tangent indicatrix of any order of a simple closed spherical curve, not a circle, exceeds the number of secondary geodesic vertices by at least four.*

For $p_k - s_k = 2 \sum_{i=1}^{k-1} s_i + g_0 \geq g_0$, and by Theorem 3.2, $g_0 \geq 4$. It is clear, by this last corollary, that a necessary condition that a closed spherical curve, C , be a tangent indicatrix of a simple closed spherical curve, not a circle, is that $p - s = l \geq 4$; that is, the curve must contain at least four geodesic inflections. Thus a figure eight curve with only two geodesic inflections cannot possibly be a tangent indicatrix of a simple closed spherical curve.

By suitably combining the relationships discussed here, it would be possible to state several interesting theorems. One illustration will suffice.

THEOREM 7.3. *If C_1 and C_2 are two mutually inverse spherical curves, with first tangent indicatrices \bar{C}_1 and \bar{C}_2 , respectively, then $\bar{p}_1 - \bar{s}_1 = \bar{p}_2 - \bar{s}_2$, where \bar{p}_i and \bar{s}_i are, respectively, the numbers of primary and secondary geodesic vertices on \bar{C}_i , $i = 1, 2$.*

For $\bar{p}_i - \bar{s}_i = g_i$, where g_i is the number of geodesic vertices on C_i , $i = 1, 2$, and by Theorem 3.1, $g_1 = g_2$.

8. Dual vertices. As a consequence of formula (7.2) there exists a very simple relationship between the geodesic vertices on the first tangent indicatrix of a closed regular space curve, C , and the dual vertices which are defined by Takasu¹⁶ as the extrema of the dual curvature, $1/P = -T/R$. By (7.2) $1/\rho_g = R/T$, where $1/\rho_g$ is the geodesic curvature of the first tangent indicatrix, whence it follows that $1/P = -\rho_g$. If $1/\rho_g \neq 0$, then $1/P$ is continuous and the dual vertices of C correspond exactly to the geodesic vertices of the tangent indicatrix. In the contrary case, however, $1/P$ becomes infinite. This occurs whenever $1/T$ becomes zero.

¹⁶ T. Takasu, *loc. cit.* 4.

If, in particular, C is a closed spherical curve, not a circle, it follows from (6.2) that $1/T$ passes through zero at least twice, and $1/P$ cannot be continuous. This fact is apparently overlooked by Takasu in proving his dual four-vertex theorem for spherical ovals. He makes use of the continuity of $1/P$ to establish the existence of at least two zeros for its derivative. The proof in question is thus invalid, but curiously the theorem itself is true, as follows at once by Corollary 7.2.2 and the relation $1/P = -\rho g$.

Takasu observes the relationship indicated above between a curve and its tangent indicatrix, but states it incorrectly as a correspondence of dual vertices of the curve and vertices of the tangent indicatrix instead of geodesic vertices. As we shall see in the next paragraph, not all vertices are geodesic vertices. As a matter of fact, the vertices of the tangent indicatrix which are not also geodesic vertices are the points that correspond to the discontinuities of $1/P$ instead of its extrema.

9. Vertices of spherical curves. In concluding this study of spherical curves, it is natural to inquire what relationship exists between the vertices and geodesic vertices of such a curve, and what can be said regarding the number of vertices. If the curve, C , lies on a sphere of radius b , then

$$\frac{1}{R^2} = \frac{1}{\rho^2} + \frac{1}{b^2}.$$

Differentiation of this equation yields

$$(9.1) \quad \frac{1}{R} \frac{d}{ds} \left(\frac{1}{R} \right) = \frac{1}{\rho} \frac{d}{ds} \left(\frac{1}{\rho} \right).$$

The points of C at which $d(1/R)/ds$, $1/\rho$, and $d(1/\rho)/ds$ change sign are, respectively, the vertices, the geodesic inflections, and the geodesic vertices. Since, for a spherical curve, $1/R \neq 0$, the left side of (9.1) changes sign precisely at the vertices. Similarly, since a geodesic inflection never coincides with a geodesic vertex, the right side of (9.1) changes sign precisely at the geodesic inflections and the geodesic vertices. This leads at once to

THEOREM 9.1. *The vertices of a spherical curve of class C''' consist of the geodesic vertices and geodesic inflections.¹⁷*

¹⁷ This theorem is true for curves of class C'' , but our argument, based on (9.1) assumes class C''' .

From this theorem it can readily be shown that vertices are not preserved by inversion, even for the special case of stereographic projection. Let a circle be tangent to an ellipse at a vertex and cut it in two other points. If a sphere is chosen on which the above circle projects stereographically into a great circle, the projection of the ellipse has geodesic inflections, for otherwise it would be a D -curve, and by Lemma 4.9 could not cross a tangent great circle. Thus the projection has at least six vertices, four geodesic vertices and at least two geodesic inflections, as compared with four vertices on the ellipse.

The statements in the last paragraph are not contrary to (2.14) for although for a plane curve $1/T \equiv 0$ and thus $d(1/R')/ds'$ is a multiple of $d(1/R)/ds$, the other factor, $(1/R)ds/ds' + 2(x|\beta)/b^2$, may become zero, whence it follows that $d(1/R')/ds'$ may change sign, even though $d(1/R)/ds \neq 0$.

Theorem 9.1 may be conveniently restated in the following form.

COROLLARY 9.1.1. *If C is any spherical curve of class C''' ,*

$$v = g + i$$

where v , g , and i are, respectively, the numbers of vertices, geodesic vertices, and geodesic inflections on C .

From this follow readily several interesting corollaries.

COROLLARY 9.1.2. *A simple closed spherical curve of class C''' , not a circle, has at least four vertices.*

For by Theorem 3.2, $g \geq 4$.

A D -curve with continuous geodesic curvature is called an oval. Using this definition, we state

COROLLARY 9.1.3. *A simple closed spherical curve of class C''' , not an oval, has at least six vertices.*

Since, by hypothesis, there are geodesic inflections, and these necessarily occur in pairs, $i \geq 2$, while by Theorem 3.2, $g \geq 4$.

COROLLARY 9.1.4. *A closed spherical curve of class C''' which contains geodesic inflections has at least four vertices.*

For by hypothesis $i \geq 2$ and $g \geq 2$ on any closed curve of class C'' .

COROLLARY 9.1.5. *A closed spherical curve of class C''' which is a tangent indicatrix of any order of a closed spherical curve, not a circle, has at least four vertices.*

By hypothesis $1/\rho \not\equiv 0$, and a necessary condition that a curve be a tangent indicatrix of a closed spherical curve is that $\int_C ds/\rho = 0$. Hence $1/\rho$ changes sign and Corollary 9.1.4 applies.

COROLLARY 9.1.6. *A closed spherical curve of class C''' which is a tangent indicatrix of any order of a simple closed spherical curve, not a circle, has at least six vertices.*

As in the last corollary $i \geq 2$, and by Corollary 7.2.2, $g \geq 4$.

THE UNIVERSITY OF WISCONSIN.

A COMPLETE CHARACTERIZATION OF SECTIONAL FAMILIES OF CURVES.*¹

By ANNETTE VASSELL.

The object of this paper is to study the geometric character of a special type of family of plane curves, the *sectional family*. A sectional family is obtained by projecting from a fixed point upon a fixed plane all the plane sections of an arbitrary surface. A set of six plane geometrical properties is found for these families and it is proved that they are characteristic. This problem was first considered by Kasner² in 1908 and the solution is analogous to his differential-geometric characterization of dynamical trajectories.³ Of the individual properties mentioned below, I is due to Kasner, as also II, III and III' for the case of developable surfaces. Moreover IV and V were suggested by his properties V and VI for dynamical trajectories.

By making use of a projective transformation in space which leaves every point of the fixed plane invariant and carries the center of projection to the point at infinity in the direction orthogonal to the plane, it is readily seen that a given sectional family obtained by central projection from one surface can always be thought of as obtained by orthogonal projection from another surface projectively related to the first. Let the fixed plane be the x, y plane, let $z = f(x, y)$ be the equation of a surface and let $z = ax + by + c$ be the equation of a general cutting plane. Projecting the plane sections orthogonally upon the x, y plane we get as the equation of the resulting family of curves

$$ax + by + c - f(x, y) = 0.$$

A sectional family is thus a certain kind of three-parameter system of plane curves.

By differentiating and eliminating the constants from the last equation we find the differential equation of the system of curves to be

$$(1) \quad (f_{xx} + 2f_{xy}y' + f_{yy}y'^2)y''' = (f_{xxx} + 3f_{xxy}y' + 3f_{xyy}y'^2 + f_{yyy}y'^3)y'' + 3(f_{xy} + f_{yy}y')y''^2.$$

* Received July 27, 1939; revised January 8, 1940.

¹ Abstract in *Bulletin of the American Mathematical Society*, vol. 45 (1939), p. 91.

² Abstracts in *Bulletin of the American Mathematical Society*, vol. 14 (1908), p. 356; vol. 36 (1930), p. 51.

³ "The Trajectories of Dynamics," *Transactions of the American Mathematical Society*, vol. 7 (1906), pp. 401-424. Also "Differential-geometric Aspects of Dynamics," *Princeton Colloquium Lectures on Mathematics* (1913; new edition 1934).

This is of the general type ⁴

$$(2) \quad y''' = G(x, y, y')y'' + H(x, y, y')y'^2.$$

Kasner proved that all triply infinite families whose differential equation is of the form (2) where G and H are any functions of x, y, y' have the following geometrical property.⁵

Property I. If to each of the ∞^1 curves having a given lineal element in common the osculating parabola is drawn at that element, the foci will lie on a circle through the point of the element.

And conversely, every system of curves possessing Property I is defined by a differential equation of the form (2). As shown in the reference, the focal circle corresponding to a lineal element x, y, y' has the equation

$$(3) \quad 2G(X^2 + Y^2) + \{3(y'^2 - 1) - y'(y'^2 + 1)H\}X \\ + \{(y'^2 + 1)H - 6y'\}Y = 0,$$

where X, Y denote current coördinates referred to axes drawn through the given point as origin and parallel to the x - and y -axes respectively.

The special form of the coefficients G and H in the sectional case indicates that sectional families possess other properties besides Property I. We observe that H has the form

$$(4) \quad H = \frac{3[y' - (w_1 + w_2)/2]}{(y' - w_1)(y' - w_2)}$$

where w_1, w_2 are the roots of the equation $f_{xx} + 2f_{xy}y' + f_{yy}y'^2 = 0$ and are therefore the projections of the asymptotic directions on the surface. We have

$$(5) \quad w_1 + w_2 = -2f_{xy}/f_{yy}, \quad w_1w_2 = f_{xx}/f_{yy}.$$

We shall show that if w_1, w_2 in (4) are any two functions of x and y , whether derived from a surface or not, the following property will hold and conversely that if this property holds then H must be of the form (4). This property was stated without proof by G. Comenetz.⁶

Property II. There exist for each point x, y of the plane two directions w_1, w_2 such that any direction y' and the reflection in y' of the tangent to the focal circle determined by x, y, y' are pairs in the involution whose fixed directions are w_1 and w_2 .

⁴ E. Kasner, "Dynamical Trajectories and the ∞^3 Plane Sections of a Surface," *Proceedings of the National Academy of Sciences*, vol. 17 (1931), pp. 370-376.

⁵ "The Trajectories of Dynamics," p. 409.

⁶ "Curvature Trajectories," *American Journal of Mathematics*, vol. 58 (1936), p. 225.

The slope of the tangent to the focal circle (3) at the given point is

$$\frac{3(y'^2 - 1) - y'(y'^2 + 1)H}{6y' - (y'^2 + 1)H}.$$

It is easily computed that the reflection in y' of this slope is $y' - 3/H$. The necessary and sufficient condition⁷ for Property II is that

$$y'(y' - 3/H) - \frac{1}{2}(w_1 + w_2)[y' + (y' - 3/H)] + w_1w_2 = 0,$$

and this equation is equivalent to (4).

When $w_1 = w_2$ (we then write them as w), the involution is singular so that the reflection in y' of the tangent to the focal circle is fixed and coincides with w .

Next we note that with the use of (5), G of (1) may be written as

$$(6) \quad G = \frac{hy'^3 + my'^2 + ny' + k}{(y' - w_1)(y' - w_2)}$$

where

$$(7) \quad h = f_{yyy}/f_{yy}, \quad m = 3f_{xyy}/f_{yy}, \quad n = 3f_{xy}/f_{yy}, \quad k = f_{xxx}/f_{yy}.$$

Moreover, when $w_1 = w_2$, it may be verified that the numerator of (6) has the factor $(y' - w)$, that is

$$(8) \quad hw^3 + mw^2 + nw + k = 0.$$

We shall prove that if h, m, n, k, w_1, w_2 are any functions of x and y subject only to the condition that if $w_1 = w_2$, (8) holds, the following property will be true of systems (2) with H given by (4), and conversely that if this property holds then G must be of the form (6) subject to (8) when $w_1 = w_2$.

Property III. In each direction through a given point O there passes one curve which has contact of third order with its circle of curvature. When the directions w_1, w_2 occurring in II are distinct, the locus of the centers of the ∞^1 hyperosculating circles, obtained by varying the initial direction, is a cubic having a rectangular node at the given point O . The nodal tangents bisect⁸ the angles made by the directions w_1, w_2 . When $w_1 = w_2$, the locus is a conic which passes through the given point in the direction w .

The condition of third order contact demands that y''' of the differential equation of circles $(1 + y'^2)y''' = 3y'y''^2$ be the same as y'' of the system (2). Equating the two expressions for y''' and then solving for y'' , we find

⁷ Graustein, "Introduction to Higher Geometry" (1930), p. 155, ex. 9.

⁸ When w_1, w_2 are the two isotropic directions, the cubic degenerates to three straight lines through the given point (if also $G \equiv 0$, it degenerates into the whole plane), and there is no rectangular node or angle bisection.

$$(9) \quad y'' = \frac{G(1 + y'^2)}{-H(1 + y'^2) + 3y'}.$$

Substituting for H from (4) and for G from (6), we have

$$(10) \quad y'' = \frac{2(hy'^3 + my'^2 + ny' + k)(1 + y'^2)}{-3[(w_1 + w_2)y'^2 - 2(w_1w_2 - 1)y' - (w_1 + w_2)]}.$$

This shows that to any x, y, y' there is one y'' . Hence to any lineal element there corresponds one curve which is hyperosculated by its circle of curvature.

The coördinates of the center of curvature for a curve at a point, with respect to that point as origin, and axes parallel to the coördinate axes, are

$$(11) \quad X = \frac{-y'(1 + y'^2)}{y''}, \quad Y = \frac{1 + y'^2}{y''}.$$

Solving for y' and y'' in (11) and substituting the values in (10) we have

$$(12) \quad hX^3 - mX^2Y + nXY^2 - kY^3 \\ - \frac{3}{2}(w_1 + w_2)X^2 - 3(w_1w_2 - 1)XY + \frac{3}{2}(w_1 + w_2)Y^2 = 0.$$

This is a cubic with a node at the given point. We observe that when we set the quadratic terms equal to zero, the product of the roots Y/X is -1 and hence the tangents at the node are perpendicular to each other. It follows from the identity

$$(w_1 + w_2)(w_1w_2) - (w_1w_2 - 1)(w_1 + w_2) - (w_1 + w_2) = 0$$

that w_1, w_2 are harmonic conjugates with respect to the roots of the quadratic, hence the roots, that is the nodal tangents, are the angle bisectors⁹ of w_1, w_2 .

When $w_1 = w_2$ an extraneous factor $X + wY$ must be removed from (12) and we obtain a conic having the w direction at the given point.

Conversely, we now ask for all systems (2) having Properties II and III.

The equation of a cubic having a rectangular node at the given point is

$$(13) \quad A_0X^3 + 3A_1X^2Y + 3A_2XY^2 + A_3Y^3 + A_4X^2 + A_5XY - A_4Y^2 = 0$$

where the coefficients A are functions of x, y . The center of a hyperosculating circle is given by (11) where y'' is defined by (9). If we substitute the center in (13), and apply Property II to the result by substituting for H from (4) and then solve for G , we find

$$(14) \quad G = \frac{-3(A_0y'^3 - 3A_1y'^2 + 3A_2y' - A_3)[(w_1 + w_2)y'^2 - 2(w_1w_2 - 1)y' - (w_1 + w_2)]}{2(y' - w_1)(y' - w_2)(A_4y'^2 - A_5y' - A_4)}.$$

⁹ Graustein, p. 155, ex. 10 and p. 152, Theorem 1. When the involution is circular, w_1, w_2 are the isotropic directions and there is no question of bisection.

This expression for G becomes simplified when we make use of the fact that the nodal tangents of the cubic bisect the angles formed by the lines $Y = w_1X$, $Y = w_2X$. The nodal tangents are defined by setting the quadratic terms of (13) equal to zero. As before we find that the condition for bisection is

$$A_4/A_5 = (w_1 + w_2)/2(w_1w_2 - 1).$$

Substituting in (14), and changing the notation somewhat we find that G simplifies to the form (6) where h, m, n, k are arbitrary in x, y .

Similarly we deal with the case $w_1 = w_2$.

We have now proved that a necessary and sufficient condition that a system of curves have an equation of the type (2) with H and G of the forms (4) and (6) respectively, that is,

$$(15) \quad (y' - w_1)(y' - w_2)y''' \\ = (hy'^3 + my'^2 + ny' + k)y'' + 3\left(y' - \frac{w_1 + w_2}{2}\right)y''^2$$

where h, m, n, k, w_1 and w_2 are arbitrary functions of x, y (except for (8) when $w_1 = w_2$) is that it possess Properties I, II, III.

Property V of Kasner's set for dynamical trajectories states a relation between the radii of curvature of the trajectory in the w direction at a given point which hyperosculates its circle of curvature, and of the line of force (the line $y' = w$) passing through the point. This suggests a similar investigation in the case of sectional families.

We shall prove the following property.

Property IV. Let the directions w_1 and w_2 of Property II be distinct. Of the curves which pass through a given point in the direction w_1 there is one which has contact of third order with its circle of curvature; the radius of curvature of this curve is $3/2$ the radius of curvature of that one of the integral curves of the direction w_1 which passes through the point. A similar statement holds for the direction w_2 . When $w_1 = w_2$, the above statement is replaced by the following: the curve in the w direction which has contact of third order with its circle of curvature has the curvature zero.¹⁰

Let $w_1 \neq w_2$. To find the radius of curvature of the curve in the w_1 direction which hyperosculates its circle of curvature we substitute in the formula for radius of curvature y'' as determined by (10) and w_1 for y' . We observe that this radius of curvature is, by definition of the cubic (12),

¹⁰ In this case the integral curve of the direction w also has curvature zero and thus the $3/2$ ratio still holds, but it is not necessary to mention this in order to have a characteristic set of properties. On the other hand merely to say that the $3/2$ ratio holds is not sufficient for a characteristic set.

identical with the segment which the normal to the w_1 direction at O intercepts on the cubic. Let us call the point of intersection of this normal with the cubic N_1 . Then

$$(16) \quad ON_1 = \frac{3(1 + w_1^2)^{3/2}(w_2 - w_1)}{2(hw_1^3 + mw_1^2 + nw_1 + k)}.$$

The radius of curvature ρ_1 of the curve $y' = w_1$ is

$$(17) \quad \rho_1 = \frac{(1 + w_1^2)^{3/2}}{w_{1x} + w_1 w_{1y}}.$$

The necessary and sufficient condition for ON_1 to be $3\rho_1/2$ is

$$(18_1) \quad hw_1^3 + mw_1^2 + nw_1 + k = (w_2 - w_1)(w_{1x} + w_1 w_{1y}).$$

Similarly the condition corresponding to the w_2 direction is

$$(18_2) \quad hw_2^3 + mw_2^2 + nw_2 + k = (w_1 - w_2)(w_{2x} + w_2 w_{2y}).$$

When $w_1 = w_2$, the denominator of (10) has the simple factor $y' - w$. In order for y'' , and hence the curvature corresponding to the w direction, to vanish when $y' = w$ the cubic in the numerator of (10) must have $(y' - w)^2$ as a factor. In view of (8), the necessary and sufficient condition for this is

$$(19) \quad 3hw^2 + 2mw + n = 0.$$

It may be verified that sectional families which have $w_1 \neq w_2$ obey (18₁) and (18₂) and those for which $w_1 = w_2$ obey (19). Hence sectional families have Property IV.

We obtain the remaining properties by differentiating and combining in various ways the relations (5) and (7). We omit the calculations and merely state the results, which can be verified from (5) and (7). The first relation found is

$$(20) \quad h_x = \left[\frac{k - (w_1 w_2)_x}{w_1 w_2} \right]_y.$$

We shall prove that (20) is the necessary and sufficient condition for the next property.

Property V. When the point O is moved, the associated cubic referred to in III changes in the following manner. Take any two fixed perpendicular directions for the x direction and the y direction; through O draw lines in these directions meeting the cubic again at A and B respectively. Also construct the normals to w_1 and w_2 at O . At A draw a line in the y direction meeting these normals in some points A' and A'' , and at B draw a line in the

x direction meeting the normals in some points B' and B'' respectively. The variation property referred to takes the form

$$(21) \quad \left[\frac{1}{AA'} + \frac{1}{AA''} \right]_x - \left[\frac{1}{BB'} + \frac{1}{BB''} \right]_y - \frac{2}{3} \left[\frac{(w_1 w_2)_x}{w_1 w_2} \right]_y = 0$$

where AA' , AA'' , BB' , BB'' are signed distances and where w_1 , w_2 denote the slopes of the directions referred to in II relative to the chosen x -direction. This is true for any pair of orthogonal directions, and therefore really expresses an intrinsic property of the system of curves.

To establish the above statement, we substitute in (20) the values of h and k from

$$(22) \quad OA = \frac{3}{2} \frac{(w_1 + w_2)}{h} \quad \text{and} \quad OB = \frac{3}{2} \frac{w_1 + w_2}{k},$$

these latter being the intercepts on the coördinate axes of the cubic (12). On simplifying the result somewhat, we find

$$\left[\frac{w_1}{OA} + \frac{w_2}{OA} \right]_x - \left[\frac{1}{w_1 OB} + \frac{1}{w_2 OB} \right]_y + \frac{2}{3} \left[\frac{(w_1 w_2)_x}{w_1 w_2} \right]_y = 0.$$

Now if we carry out the construction as expressed in V we find from triangles $AA'O$ and $BB'O$ that $-OA = w_1 AA'$ and $-BB' = w_1 OB$, and from triangles $AA''O$ and $BB''O$ that $-OA = w_2 AA''$ and $-BB'' = w_2 OB$. On substituting these in the above equation, we obtain (21).

The final property depends on the following two relations.

$$(23) \quad 2 \left(\frac{w_1 w_2}{w_1 + w_2} \right)^2 h + \frac{k}{w_1 + w_2} = \left[\frac{w_1 w_2}{w_1 + w_2} \right]_x - \frac{2 w_1 w_2 (w_1 w_2)_y}{(w_1 + w_2)^2},$$

$$(24) \quad (w_1 w_2)(w_1 + w_2)h + 2k = 2(w_1 w_2)_x - (w_1 w_2)(w_1 + w_2)_y.$$

If we now use the relations (22) in (23) and (24) we obtain

$$(25) \quad \frac{2(w_1 w_2)^2}{OA} + \frac{w_1 + w_2}{OB} + \frac{4}{3} \frac{(w_1 w_2)(w_1 w_2)_y}{w_1 + w_2} - \frac{2}{3}(w_1 + w_2) \left[\frac{w_1 w_2}{w_1 + w_2} \right]_x = 0,$$

$$(26) \quad \frac{(w_1 w_2)(w_1 + w_2)^2}{OA} + 2 \frac{(w_1 + w_2)}{OB} - \frac{4}{3}(w_1 w_2)_x + \frac{2}{3}(w_1 w_2)(w_1 + w_2)_y = 0.$$

Property VI. Let the intercepts OA and OB be constructed as in V. The variation of the directions w_1 and w_2 as the point O is moved is related to these intercepts by the equations (25) and (26).

Obviously (23) and (24) are the necessary and sufficient conditions for Property VI.

We shall now prove that Properties I-VI are sufficient for sectional families. We have already seen that any system of curves having Properties I, II and III will have an equation of the form (15). Our problem is to show that if we apply (18₁), (18₂) (or if $w_1 = w_2$, (8), (19)), (21), (25) and (26) to a system of curves (15), the curves will be the projections of the plane sections of a surface; that is, that w_1, w_2, h, m, n, k will be of the forms (5) and (7), where f is some function of x and y . Instead of (21), (25) and (26) we may use (20), (23), and (24), which are equivalent to them.

Equation (20) is an integrability condition; hence a function $F(x, y)$ exists which satisfies the equations

$$(27) \quad F_y = h, \quad F_x = \frac{k - (w_1 w_2)_x}{w_1 w_2}.$$

If we place the expressions for h and k from (27) in (23) and multiply the result by e^F we have

$$(e^F w_1 w_2)_y = -\frac{1}{2} \{e^F (w_1 + w_2)\}_x.$$

Therefore a function H exists such that

$$(28) \quad e^F w_1 w_2 = H_x \quad \text{and} \quad -\frac{1}{2} e^F (w_1 + w_2) = H_y.$$

Now if we substitute (27) and (28) in (24) so as to eliminate h, k, w_1 and w_2 , and simplify, we find that $H_{yy} = (e^F)_x$. This equation means that a function g exists for which $H_y = g_x$ and $e^F = g_y$. The first of these two equations further says that a function f exists for which $H = f_x$ and $g = f_y$. We have then

$$(29) \quad e^F = f_{yy} \quad \text{and} \quad H = f_x.$$

Consequently w_1, w_2 obey (5). Next we substitute (5) and (29) in the equations (18₁), (18₂) (or if $w_1 = w_2$, in (8), (19)) and (27) and solve them for h, m, n, k . We find that they satisfy (7).

We have now proved that *every sectional family possesses Properties I-VI and every family of curves possessing Properties I-VI is a sectional family.*

Developable surfaces. On a developable surface the asymptotic directions coincide, hence $w_1 = w_2$. Therefore a sectional family derived from a developable surface, which we may call a developable system, has Properties I-VI in the simpler form to which they reduce when $w_1 = w_2$.

Conversely, every family of curves having Properties I-VI in the reduced form is a developable system. Thus the reduced set of six properties is a characteristic set for developable systems. In this case G in (2) is linear in y' and H is $3/(y' - w)$.

When w_1 is set equal to w_2 , (21) in Property V becomes

$$(30) \quad \frac{\partial}{\partial x} \left(\frac{1}{AA'} \right) - \frac{\partial}{\partial y} \left(\frac{1}{BB'} \right) + \frac{2}{3} \left(\frac{w_x}{w} \right)_y = 0$$

and Property II says that y' bisects the angle between w and the reflection in y' of the tangent to the focal circle. For developable systems one asymptote of the conic in Property III is perpendicular to the w direction.

Now a given type of family may be characterized by more than one set of properties. In the case of developable systems it is possible to replace VI by the following more elegant pair of properties.

(A) The integral curves of the direction w are straight lines.

(B) Let α be the angle between the asymptotes to the conic of III, let K be the curvature of the conic at the point O , and let K_1 be the curvature at O of the orthogonal trajectories to the straight lines $y' = w$. Then

$$K \sin \alpha + 2K_1 \cos \alpha = 0.$$

Property (A) is well known although it has not been stated in this connection before. The corresponding analytic statement is $w_x + ww_y = 0$. The $(w_x/w)_y$ in (30) then becomes $-w_{yy}$.

A property of sectional families derived from non-developable surfaces which is analogous to (A) may be obtained from (5) by differentiating, and eliminating f and its derivatives. The eliminant is found to be

$$(w_1 - w_2)(w_{1xx} + w_{2xx} + 2w_1w_{2xy} + 2w_2w_{1xy} + w_1^2w_{2yy} + w_2^2w_{1yy}) \\ + 2w_1^2w_{2y}^2 - 2w_2^2w_{1y}^2 + 4w_1w_{2x}w_{2y} - 4w_2w_{1x}w_{1y} + 2w_{2x}^2 - 2w_{1x}^2 = 0.$$

This has the following geometric significance. Let θ be the angle between the directions w_1 and w_2 , let γ and Γ be the curvatures and let s and S be the lengths of arc along the curves $y' = w_1$ and $y' = w_2$ respectively; then

$$\sin \theta \left(\frac{d^2\theta}{dS^2} - \frac{d\Gamma}{dS} \right) - \cos \theta \left(\Gamma - \frac{d\theta}{dS} \right) \left(\Gamma - 2 \frac{d\theta}{dS} \right) \\ = \sin \theta \left(\frac{d^2\theta}{ds^2} + \frac{d\gamma}{ds} \right) - \cos \theta \left(\gamma + \frac{d\theta}{ds} \right) \left(\gamma + 2 \frac{d\theta}{ds} \right).$$

This relation is a new characterization of asymptotic nets.¹¹

An alternative for Property III for developable systems is:

¹¹ References to other characterizations of asymptotic nets may be found in Fubini and Čech, "Introduction à la Géométrie Projective Différentielle des Surfaces" (1931), p. 191.

Property III'. The locus of the centers of the ∞^1 circles corresponding to the elements at a given point is a conic with that point as focus.

For a given developable surface the conic is

$$\left(\frac{f_{xxx}}{\sqrt{f_{xx}}} + \frac{f_{yyy}}{\sqrt{f_{yy}}} \right) \sqrt{X^2 + Y^2} = \sqrt{f_{xx} + f_{yy}} \left(\frac{f_{xxx}}{f_{xx}} X + \frac{f_{yyy}}{f_{yy}} Y + \frac{3}{2} \right).$$

The analogue of III' for families derived from non-developable surfaces has not been worked out completely but it is certain that the locus of the centers of the focal circles at a point is a cubic curve. The coefficients are very long expressions in terms of $f_{xx}, f_{xy}, \dots, f_{yyy}$ and they are symmetrical in the subscripts x and y . The constant term turns out to be $M(L^2 - 4M)^2$ where M is the Hessian and L the Laplacian of the surface.

Ruled surfaces. Ruled surfaces may be characterized analytically by the fact that the cubic in y' in the numerator and the quadratic in the denominator of G (that is, the coefficients of y'' and y''' in (1)), have a linear factor in common. For the condition for the existence of such a factor is exactly the differential equation of ruled surfaces found by Monge.

In order to give a characteristic set of geometrical properties for sectional families derived from ruled surfaces we have only to add the following property to the previous set.

Property VII. One of the two families of integral curves of the directions w_1, w_2 consists of straight lines.

This is so because a ruled surface is a surface one of whose two families of asymptotic lines are straight lines. In this case the family of straight lines is at the same time part of the sectional family since it is the projection of the rulings on the surface. These straight lines belong to the hyperosculated curves referred to in Property IV.

The condition for VII is

$$(w_{1x} + w_1 w_{1y})(w_{2x} + w_2 w_{2y}) = 0,$$

where w_1, w_2 are given by (5). Incidentally it follows that this condition is equivalent to Monge's equation for ruled surfaces.

EXACTLY $(k, 1)$ TRANSFORMATIONS ON CONNECTED LINEAR GRAPHS.*

By O. G. HARROLD, JR.¹

1. Introduction. In a paper by G. T. Whyburn [1]² there is given, among other results, a detailed study of the behavior of interior transformations on linear graphs. The results suggest a connection between these transformations and transformations which are exactly $(k, 1)$.³ This paper gives a more or less detailed account of exactly $(k, 1)$ continuous transformations defined on a connected linear graph. Since we are considering a connected graph, this type of transformation includes all local homeomorphisms defined on the given set [2].

In 2. results are given concerning exactly $(k, 1)$ mappings defined on Peano spaces of varying degrees of complexity. It is hoped that several of these results will be of use in attacking the problem of determining precisely what topological structure a set must have in order that an exactly $(k, 1)$ continuous mapping can be defined on it for $k > 1$ [3].⁴ In 3. an example is given showing that an exactly $(3, 1)$ image of a graph need not be a graph. In 4. the case $k = 2$ is given special attention. It is shown that an exactly $(2, 1)$ image of a graph A is a graph B , and furthermore, there exist subdivisions of A and B into finite complexes K_A and K_B , respectively, such that the transformation of K_A into K_B is simplicial. A formula is given which not only relates the structure of A to B but also actually limits the type of sets A on which an exactly $(2, 1)$ mapping can be defined.

2. Exactly $(k, 1)$ transformations in general.

2.1. Let $f(A) = B$, where A and B are arcs. If f is at most $(k, 1)$ on A , there is an open set U dense on A such that f is topological on each component of U .

* Received March 7, 1940.

¹ National Research Fellow.

² The numerals in brackets refer to the bibliography at the end of the paper.

³ All transformations considered in this paper are single valued and continuous. By an exactly $(k, 1)$ transformation is meant a continuous mapping such that each point of the image space has exactly k inverse points. By a $(k, 1)$ mapping is meant a continuous mapping such that every point of the image set has at most k inverses.

⁴ See also an abstract by J. H. Roberts in the *Bulletin of the American Mathematical Society*, 45-11-433.

Proof. It is supposed that A and B are non-degenerate arcs. The assertion is true for $k = 1$. For $k > 1$ it will suffice to show that on any sub-arc of A there is an open set U on which f is a homeomorphism. Denote the end-points of A by a^1 and a^2 and those of B by b^1 and b^2 . It may be supposed that A is a sub-arc of the given arc A such that $f^{-1}(b^1) = a^1$ and $f^{-1}(b^2) = a^2$. The inductive assertion is that the statement is true for $k - 1$. If every point in the interior of B has at most $k - 1$ inverses in A , the desired open set U exists by our hypothesis. If $x \in B - (b^1 + b^2)$ has k inverses in A , let the first and last of these in order from a^1 to a^2 be x^1 and x^2 , respectively. There is an open interval V in B with x as an end-point such that for every $y \in V$, $f^{-1}(y)(a^1x^1 + x^2a^2)$ has at most $k - 2$ points, otherwise, since $f(x^1x^2)$ is a non-degenerate arc in B , some point in B would have at least $k + 1$ inverses. For some z on a^1x^1 (or x^2a^2) the open interval W between z and x^1 (or x^2) maps into a subset of V . By the choice of V , f must be $(k - 1, 1)$ on \bar{W} , hence by the inductive assumption, there is an open set U dense in \bar{W} such that f is topological on each component of U . This proves 2.1.

2.2. If f is exactly $(k, 1)$ on the stably regular curve A ,⁵ $B = f(A)$ is stably regular.

Proof. First, B can contain no non-degenerate continuum X which contains no arc. For this would imply that $f^{-1}(X)$ is totally disconnected, which is not possible for an exactly $(k, 1)$ mapping.⁶ Thus, assuming that B can contain a non-degenerate continuum X such that $X \subset \overline{B - X}$, we may take X to be an arc. It will be shown that this denial that B is stably regular leads to a contradiction. Since $f^{-1}(X)$ is not totally disconnected, there is a non-degenerate continuum $Y \subset f^{-1}(X)$. Since A is stably regular, we may take Y to be a free arc⁷ in A . Then $f(Y) = X^1 \subset X$. Hence we may apply 2.1 and further restrict Y to be an arc mapping topologically into some arc X^1 in X . Clearly, $X^1 \subset \overline{B - X^1}$. Each point z in the interior of X^1 has exactly $k - 1$ inverses in $\overline{A - Y}$. Thus any arc X_1^1 wholly in the interior of X^1 has an inverse $f^{-1}(X_1^1)(\overline{A - Y})$ which is not totally disconnected, since f is exactly $(k - 1, 1)$ on the compact set $f^{-1}(X_1^1)(\overline{A - Y})$. Hence there is a free arc $Z \subset \overline{A - Y}$ such that f maps Z topologically into $X^2 \subset X^1$. As before, $X^2 \subset \overline{B - X^2}$. After a finite number of steps there are determined k free arcs Y, Z, \dots, W in A such that each contains an arc T_Y, T_Z, \dots, T_W

⁵ A continuum M is said to be stably regular (beständig regular) provided that for no non-degenerate continuum T does $T \subset \overline{M - T}$.

⁶ See [3] and the references given therein.

⁷ The arc T is said to be free in X provided no interior point of T is a limit point of $X - T$. It is essential to notice that a free arc is a closed point set.

mapping topologically onto X^0 , where X^0 is a non-degenerate arc such that $X^0 \subset \overline{B - X^0}$. Let x be any interior point of X^0 . Let $x_n \rightarrow x$, $x_n \in B - X$. Since $f^{-1}(x_n)$ and $(Y + Z + \dots + W)$ have no common points, this implies that the compact set $A - (T_Y + T_Z + \dots + T_W)$ has an inverse to x , hence x has $k + 1$ inverses in all, which is not possible. This permits us to state also

2.3. *If f is exactly $(k, 1)$ on the stably regular curve A , to each non-degenerate arc G in $B = f(A)$ there is a non-degenerate sub-arc G^1 of G such that $f^{-1}(G^1)$ consists of k arcs each mapping topologically onto G^1 .*

2.4. *If f is exactly $(k, 1)$ on the Peano space A and x is an end-point of $B = f(A)$, each point of $f^{-1}(x)$ is an end-point of A .*

Proof. Denote the Urysohn-Menger order of x by $o(x)$. Set $f^{-1}(x) = x^1 + x^2 + \dots + x^k$. Let (T_j^i) , $j = 1, 2, \dots, n_i$, be a set of n_i arcs in A each terminating at x^i , but with no other common point (by pairs). Suppose each set chosen so that $\Sigma T_j^i \cdot T_{j'}^{i'} = 0$, $i \neq i'$. Then $C = \prod_{i=1}^k \prod_{j=1}^{n_i} f(T_j^i)$ contains an infinite sequence of distinct points converging to x since $o(x) = 1$. Clearly, each point of $C - x$ has $\sum_{i=1}^k n_i = k$ inverses, hence each $n_i = 1$ and each $o(x^i) = 1$.

If $V_\epsilon(x)$ denotes a region in B of diameter less than ϵ , and if $V_\epsilon(x) - x$ has $o(x)$ components for ϵ sufficiently small, then x is an end-point of each such component. We have

2.5. *If f is exactly $(k, 1)$ on the Peano space A and $x \in B = f(A)$ is such that each sufficiently small region in B containing x is cut into $o(x)$ components by the removal of x , then $\sum_{i=1}^k o(x^i) \leq o(x) \cdot k$.*

2.6. *Let f be exactly $(k, 1)$ on the stably regular curve A . If X is a free arc in $B = f(A)$ containing no point of $E = f(E^0)$, where E^0 is the set of branch points plus end-points of A , then $f^{-1}(X)$ has at most $2k - 1$ components.*

Proof. Let X be a free arc in $B - E$. Then $f^{-1}(X) \subset A - E^0$. Since f is exactly $(k, 1)$ on A , $f^{-1}(X)$ is not totally disconnected. Let J be a non-degenerate component of $f^{-1}(X)$. By the choice of E , J is necessarily an arc (a free arc).^{*} By the continuity of f , each end-point of J must map into an end-point of X (not necessarily the same end-point). There remain at most

^{*}It is evident that J could be only a simple closed curve or arc, since all of its points are of order two. The first possibility is ruled out by 2.7 (which does not depend on 2.6).

$2k - 2$ points in $f^{-1}(X) - J$ to be located which map into an end-point of X . Clearly, any isolated point of $f^{-1}(X)$ must map, by continuity, into an end-point of X . Hence each component of $f^{-1}(X) - J$ contains at least one point which is carried into an end-point. Thus $f^{-1}(X) - J$ can have at most $2k - 2$ components.

2.7. *If f is exactly $(k, 1)$, $k > 1$, on the continuum A , $B = f(A)$ is not an arc.*

Proof. Suppose, on the contrary, that B is an arc for some exactly $(k, 1)$, $k > 1$, mapping defined on a continuum A . Now it is known that if a regular curve (Menger) is obtained from a continuum A by an at most $(k, 1)$ continuous mapping, then A is likewise regular [4]. Thus A will be assumed to be regular (hereditary local connectedness is sufficient). We take B to be the unit interval $0 \leq y \leq 1$. First, there is a proper subcontinuum A^1 of A such that f is exactly $(k, 1)$ on A^1 and $f(A^1) = B^1$, where B^1 is the interval $0 \leq y \leq b^1 < 1$. By 2.4, each point of $f^{-1}(1) = x^1 + x^2 \cdots + x^k$ is an end-point of A . Let T^1, T^2, \dots, T^k be k arcs in A containing x^1, \dots, x^k , respectively, and such that $T^i T^j = 0$, $i \neq j$. The set $C = \prod_1^k f(T^i)$ is an arc in B containing $y = 1$. Setting $S^i = T^i f^{-1}(C)$ and noting that f is $(1, 1)$ on S^i , S^i is an arc. In fact, since B is an arc, the arcs S^i are free in A . Let the end-points of C be z and $y = 1$. For any $z < b^1 < 1$, the set $E(0 \leq y \leq b^1)$ is A minus k open free arcs (each containing an end-point of A) which is a continuum A^1 on which f is exactly $(k, 1)$.

Next, the property of being a continuum in A on which f is exactly $(k, 1)$ is an inducible property, for if $A^0 = \prod_1^\infty A^i$, where each A^i is a continuum in A on which f is exactly $(k, 1)$ and $A^{i+1} \subset A^i$, then $z \in A^0$ implies $f^{-1}f(z) \subset A^0$. It follows that there is a continuum A^0 in A which is irreducible with respect to the property of being a continuum containing $f^{-1}(0)$ and on which f is exactly $(k, 1)$. If $f(A^0)$ is non-degenerate, we get a contradiction, for, A^0 is a Peano space (A is hereditarily locally connected) and the first remark will apply, hence A^0 is not irreducible. If $f(A^0) = 0$, we again get a contradiction, for A^0 is a continuum containing only $x^1 + x^2 \cdots + x^k = f^{-1}(0)$.

2.8. *If f is exactly $(k, 1)$ on the compact, hereditarily locally connected space A and $B = f(A)$ is an arc, there is an arc B^1 in B such that $f^{-1}(B^1)$ consists of k arcs each mapping topologically onto B^1 .*

Proof. The statement is trivial if B reduces to a single point. Since f is exactly $(k, 1)$, there exists a non-degenerate continuum $A^1 \subset A$. Since

A is hereditarily locally connected, A^1 may be taken to be an arc. Hence, by 2.1, there is an arc A_1^1 mapped topologically by f into $B_1^1 \subset B$. Since f is exactly $(k-1, 1)$ on the half-compact space $A - A_1^1$, there is a non-degenerate continuum $A^2 \subset f^{-1}(B_1^1) \cdot (A - A_1^1)$. As before, we take A^2 to be an arc, and applying 2.1, there is an arc $A_2^2 \subset A^2$ on which f is topological. Set $f(A_2^2) = B_2^1$. Then A_2^2 and A_1^1 each contain an arc mapping topologically onto B_2^1 . After k such steps, we obtain arcs $A_1^1, A_2^2, \dots, A_k^k$ each of which contain an arc $C^i = A_i^i f^{-1}(B_k^1)$, $i = 1, 2, \dots, k$, mapping topologically onto B_k^1 .

2.9. If f is exactly $(k, 1)$ on the continuum A and $B = f(A)$ is stably regular, so also is A .

Proof. Case 1. The continuum A is hereditarily locally connected. Suppose T is a continuum of condensation of A . Since A is hereditarily locally connected, we may take T to be an arc. Let $S = f(T)$. Let Y be a free arc in S . Since Y is free, $f^{-1}(Y)$ can have only a finite number of components and is thus hereditarily locally connected. Applying 2.8 to $f^{-1}(Y)$, there exists a set of k arcs T^1, T^2, \dots, T^k (mutually separated by pairs) in A such that each T^i maps homeomorphically onto $X \subset Y$. Since the sum of k mutually separated arcs cannot contain a continuum of condensation, there are points in any neighborhood of a point $x \in T$ which belong to $A - (T + \sum_{i=1}^k T^i)$. Let z be any interior point of the arc X . Since $f(T) \supset X$, $f^{-1}(z)T \neq \emptyset$. Let $x^0 \in T \cdot f^{-1}(z)$. Let $x_n \rightarrow x^0$, $x_n \in A - (\sum_{i=1}^k T^i + T)$. Then by continuity, $f(x_n) \rightarrow f(x)$. But $f(x)$ is an interior point of a free arc X all of whose inverses have been located in ΣT^i , which is a contradiction.

Case 2. The continuum A contains a continuum of convergence. The transformation f being at most $(k, 1)$ and A irregular (in the sense of Menger), B is likewise [4].

3. The original set A is a graph. If A is a connected linear graph, the image set $B = f(A)$ under an exactly $(k, 1)$ mapping is a stably regular curve which has at most a finite number of end-points. The following example shows that an exactly $(3, 1)$ continuous mapping defined on a graph need not give a graph.

EXAMPLE. Two basic mappings of an arc into an arc will be defined. Let C be the interval $0 \leq t \leq 1$. Let D be the interval $0 \leq y \leq 1$. Let the closed interval of $C(D)$ between $(n+1)^{-1}$ and $(n)^{-1}$ be denoted by $C_n(D_n)$. Define f as follows: Map C_1 topologically into D_1 with $f(1) = 1$. Map C_2

(topologically) into D_1 with $f(1/2) = 1/2$. Map C_3 into $D_1 + D_2$ with $f(1/3) = 1$. In general, C_{2n} is mapped onto D_n such that as t decreases y increases and C_{2n-1} is mapped onto $D_{n-1} + D_n$ such that as t decreases y decreases. Finally, $f(0) = 0$. Each point in the interior of D has three inverses in C , while $y = 0$ has one and $y = 1$ has two inverses. This will be called a mapping of type (α) . By demanding that the point which generates D oscillate in the same manner near $y = 1$ as it does near $y = 0$, a continuous mapping of C on D is effected which has the same properties as the one above except that both $y = 1$ and $y = 0$ now have precisely one inverse. This will be referred to as a mapping of type (β) .

Let X^1, X^2 and X^3 be the intervals $0 \leq x \leq 1, y = 1, 2, 3$, respectively. Let J denote $x = 0, 1 \leq y \leq 3$. Set $A = J + X^1 + X^2 + X^3$. Let X_n^i be the sub-interval of X^i between $x = 1/n$ and $x = 1/(n+1)$. For a fixed n define on $X_n^i, i = 1, 2, 3$ a mapping of type (β) . Then identify the end-points of the image arcs corresponding to $x = 1/n$ and identify the end-points of the image arcs corresponding to $x = 1/(n+1)$. Thus $f(X_n^1 + X_n^2 + X_n^3) = Y_n$ is a theta curve (i. e. of the form of the letter θ). Every point of Y_n has exactly three inverses on $X_n^1 + X_n^2 + X_n^3$. Repeating this for each $n = 1, 2, 3, \dots$ and setting $Y = \overline{\sum Y_n}$, we obtain a continuous exactly $(3, 1)$ mapping of $X^1 + X^2 + X^3$ onto Y . Evidently Y is merely the enclosure of a sequence of theta curves converging down to a single point such that $Y_n \cdot Y_{n+1}$ is a single point. Now on J define a mapping similar to type (α) such that the points of JX^1, JX^2 and JX^3 are identified. Thus $f(J) = J^1$ is a topological circle. Setting $B = f(A)$, we have an exactly $(3, 1)$ continuous mapping of the triod A onto B . The image B is clearly no graph, since it contains an infinite sequence of simple closed curves.

A continuous transformation f is said to be *locally interior* at the point x provided that $f(x)$ is not a boundary point of the transform of any open set U containing x . The above mappings of type (α) and (β) fail to have the property of being locally interior at the points $t = 1/n$, hence the above exactly $(3, 1)$ mapping is not locally interior at infinitely many points. It follows from 2.3 that any exactly $(k, 1)$ mapping defined on a stably regular continuum is locally interior except perhaps for a closed set of dimension zero.

3.1. If f is an exactly $(k, 1), k > 1$, mapping defined on the graph $A, B = f(A)$ contains a simple closed curve.

Proof. It is to be shown that B is not a dendrite. From 2.4 and the fact that A is a graph it follows that B has at most a finite number of end-points. Let B have n end-points. The assertion has been proved for $n = 2$

(2.7). Assuming the statement true for a dendritic graph with n end-points, it will be shown to be true for $n + 1$. Let x be an end-point of B . Denote the maximal free arc containing x by X . Set $C = \overline{B - X}$. Set $D = f^{-1}(C)$. The property of being a continuum in A which contains D and on which f is exactly $(k, 1)$ is inducible. Hence there is a continuum A^0 in A which is irreducible with respect to this property. If $f(A^0)$ contains C as a proper subset, by precisely the same reduction as was made in the proof of 2.7 we can find a subcontinuum H in A^0 such that $C \subset f(H)$, $f(H)$ is a proper subset of $f(A^0)$ and f is exactly $(k, 1)$ on H . But this denies the irreducibility property of A^0 . Hence $f(A^0) = C$. But C has one less end-point than B , hence, by the inductive hypothesis, f is not exactly $(k, 1)$ on A^0 .

The preceding results, in so far as they apply to the case in which A is a graph, may be summarized as follows.

3.2. *Let f be a continuous exactly $(k, 1)$ transformation defined on the connected linear graph A . The image $B = f(A)$ is a stably regular curve. The curve B is never a dendrite, and for $k > 2$ need not be a graph. The function f is locally interior at all points of A except possibly for a closed set of dimension 0. There is a closed set D of dimension 0 in B such that each component of $B - D$ is an open free arc whose inverse is precisely k open free arcs in A each mapping topologically onto the common image. Each free arc X in B containing no point of $E = f(E^0)$, where E^0 is the set of end-points plus branch points of A , is such that $f^{-1}(X)$ has at most $2k - 1$ components.*

While the above theorem fails to give an exact statement of what the image B will be in terms of the properties of A , it does show that an exactly $(k, 1)$ mapping on a graph has some of the characteristics of an interior mapping. For instance, it produces only a slightly more complicated curve than the original set. This is meant only in a relative sense, of course. It is known that an at most $(3, 1)$ mapping on an arc can increase dimensionality, while a $(2, 1)$ mapping on an arc can produce a curve containing a continuum of convergence. (For interior transformations the property of being a graph is preserved).

4. Exactly $(2, 1)$ transformations on a connected linear graph. The results in this case are much more precise, as would be expected, of course. The underlying reason for this, actually, seems to be that 2.1 can be strengthened to read

4.1. *If f maps the arc A into the arc B and is at most $(2, 1)$ on A , then f is topological on A provided it preserves end-points.⁹*

⁹ See [3], Lemma A.

As intermediate conclusions to the main results of this section we show

4.2. If f is exactly $(2, 1)$ on the graph A , then (i) there exists at most a finite number of points x in $B = f(A)$ such that x is the vertex of a triod in B containing free arcs xy and xz ; (ii) all but a finite number of the maximal free arcs X in B are such that $f^{-1}(X)$ has exactly two inverse components; (iii) no point x in B is the vertex of infinitely many free arcs.

Proof. (i) Since the set E^0 of end-points and branch points in A is finite, $f(E^0)$ is a finite set. Hence if there were infinitely many points x in B with the asserted property, there would be one, say x , such that $x \in f(E^0)$. Let T be the enclosure of a region in B containing a triod with x as vertex and such that two of the arcs of this triod, say xy and xz , are free arcs. It is supposed further that $Tf(E^0) = 0$. Set $f^{-1}(x) = x^1 + x^2$. The points x^1 and x^2 are in the interior of free arcs in A . Let X^{ij} , $i, j = 1, 2$ be free arcs in A having only the point x^i , $j = 1, 2$ in common and such that $f(X^{ij}) \subset T$. It may be supposed that $(X^{11} + X^{12})(X^{21} + X^{22}) = 0$. Denote three components of $T - x$ by xy , xz and W . Suppose $f^{-1}(xy) \cdot X^{11} \neq 0$. Let $p \in f^{-1}(xy) \cdot X^{11}$. Then $p \neq x^1$. Denote the subarc of X^{11} from p to x^1 by px^1 . Let the last point on px^1 in $f^{-1}f(p)$ be p^1 . Then f is topological on p^1x^1 by 4.1. Since f is exactly $(2, 1)$, $f^{-1}(xy) \cdot X^{12}$ (say) $\neq 0$. Similarly, there is an arc p^2x^1 in X^{12} on which f is topological. Hence there must be four arcs p^1x^1 , p^2x^1 , q^1x^2 and q^2x^2 on which f is topological and such that $f(p^1x^1 + p^2x^1 + q^1x^2 + q^2x^2) \subset xy + xz$. Further, the sum of these four arcs contains an open set containing $x^1 + x^2$. The continuity of f , however, implies that x have at least one more inverse, since W has x as a limit point. This denies the $(2, 1)$ property. The property (iii) follows from an analogous argument.

(ii). The maximal free arcs are uniquely determined. Suppose there are infinitely many of them. There is one, call it K , such that each point of $f^{-1}(K)$ is of Urysohn-Menger order two and contains no topological circle of A . Since f is exactly $(2, 1)$, we have by 2.6 that $f^{-1}(K)$ has at most 3 components. Since f is exactly $(2, 1)$, at least one of these components must be non-degenerate. By the choice of K it must be an arc. Let T be such a non-degenerate component of $f^{-1}(K)$. Since K is a free arc, end-points of T must map into end-points of K . *Case 1.* If the end-points of K are contained in the image set of the end-points of T , then by 4.1, f must be topological on T . Since f is $(1, 1)$ on the set $f^{-1}(K) - T$, which has at most 2 components, $f^{-1}(K) - T$ consists of a single component T^1 which maps topologically onto K . *Case 2.* If one end-point of K contains the image set of the end-points of T , we distinguish two cases a) and b), according as $f(T)$ contains the other end-point of K or not. In the first mentioned possibility it is easy to

show that T is the sum of two arcs T^1 and T^2 having only a common end-point and such that each $f(T^i) = K$. Hence all points of K have two inverses on T except one end-point. Thus $f^{-1}(K) - T$ consists of a single point, and $f^{-1}(K)$ has two components. In the second mentioned possibility, $f(T) = K^1$ is a subarc of K . In this case T contains two inverses to all points of K^1 except the end-point of K^1 in the interior of K . Set $K^2 = \overline{K - K^1}$. Since K^2 is free, $f^{-1}(K^2) = T^0$ can only have a single non-degenerate component, which is seen to be T^0 itself. Each point in the interior of K^2 has two inverses in T^0 . The point $K^1 \cdot K^2$ has one inverse in T and one in T^0 . Thus any arc K in B of the type we have described has exactly two inverse components, i. e. $f^{-1}(K)$ has two components.

4.3. *Let f be exactly $(2, 1)$ on the connected linear graph A . Then $B = f(A)$ is a graph. There exists subdivisions of A and B into finite complexes K_A and K_B such that the transformation of K_A into K_B is simplicial.*

Proof. First, B is a graph. To this end an upper semi-continuous decomposition of B is effected. From 2.2, there are free arcs in every open set in B . Any free arc is contained in either a maximal free arc or in a simple closed curve having just one point in common with the rest of B . (If B is a simple closed curve, our conclusion is already attained). To each free arc T in B there is a connected set Γ containing T which is a sum of such simple closed curves (of the type just mentioned) and maximal free arcs and which is maximal in this regard. By 4.2 (i), (iii), Γ contains only a finite number of simple closed curves or maximal free arcs. Hence Γ is a graph. Now the elements of the decomposition are to be (1) the graphs Γ which contain a simple closed curve or two maximal free arcs, (2) the maximal free arcs in B (not already in (1)) which have more than two inverse components, and, (3) the points in B not in an element of type (1) or (2). It will be shown that there is only one element in this decomposition. If there is only one element, clearly it must be of type (1). Next, the above definitions do give an upper semi-continuous decomposition. First, the elements are disjoint. From 4.2 (i), the elements of type (1) and hence all are closed. Also, by 4.2, there are only a finite number of elements of type (1) or (2) so clearly this gives an upper semi-continuous decomposition. Let C be an image space of a corresponding continuous transformation g defined on B , $g(B) = C$. Evidently, C can have no continuum of condensation, thus C is stably regular. If C contains more than a single point, the inverse of a point in C with a finite number of exceptions is a single point. Denote this exceptional set by $E \subset C$. By the manner of selection of the elements of the decomposition,

no two maximal free arcs in C can intersect. Also, C can contain no simple closed curve having only one point in common with the rest of C . The curve C has only a finite number of end-points since B has only a finite number. (The function g is topological on $B - g^{-1}(E)$). Hence we are in a position to define another upper semi-continuous decomposition, this one taking place on C . The elements of this decomposition are defined to be the maximal free arcs in C and the points in no free arc. It is known that this decomposition defined on such a curve C gives a hyperspace D containing no free arc [3]. Setting $h(C) = D$, $D = hgf(A)$. The continuous transformation of A into D can be factored into a monotone transformation f_1 followed by a light transformation f_2 , where $f_1(A) = A^1$ and $f_2(A^1) = D$. This factorization may be so accomplished that the 'points' in A^1 are the components of inverse sets to the mapping $hgf(A) = D$ [5]. Since f_1 is monotone and A is a graph, A^1 is a graph.¹⁰ The set $h(E)$ contains only a finite number of points. Let x be any point of $D - h(E)$. It is readily seen that either $f^{-1}g^{-1}h^{-1}(x)$ consists of exactly two arcs (one of which may be degenerate) or two points in A (according as $h^{-1}(x)$ is an arc or a point), hence $f_2^{-1}(x)$ consists of *exactly* two points. Thus the transformation f_2 carrying the graph A^1 into D is exactly $(2, 1)$ except for the points of $h(E)$. Since D contains no free arc, there is a non-degenerate arc T in $D - h(E)$ such that $T \subset \overline{D - h(E)}$. Now precisely as in the proof of 2.2 (taking $A = A^1$), this leads to a contradiction. Hence C is a single point, i. e. B is a graph.

It will now be shown that there exist subdivisions of A and B into finite complexes K_A and K_B , respectively, such that the transformation of K_A into K_B is simplicial. Let E be a finite set in B such that it contains all points of B of order $\neq 2$ and such that each component of $B - E$ has an enclosure which is an arc uniquely determined by its end-points. Set $E^0 = f^{-1}(E)$. Add to E^0 a finite set $F = f^{-1}f(F)$ such that each component of $A - (E^0 + F)$ is an open free arc (whose enclosure is an *arc*) uniquely determined by its end-points. Consider any component U of $B - (E + f(F))$. Since each non-degenerate component of $f^{-1}(\bar{U})$ contains only points of order ≤ 2 and contains no simple closed curve, the reasoning in the proof of 4.2 (ii) can be applied, hence $f^{-1}(\bar{U})$ has two components. If $\bar{U} = K$ gives rise to Case 1, $f^{-1}(\bar{U})$ consists of two disjoint arcs which are mapped topologically onto \bar{U} . By definition of E and F , these arcs are edges of the complex introduced into A by the points $F + E^0$. If \bar{U} gives Case 2a, $f^{-1}(\bar{U})$ has a single non-degenerate inverse component which is the sum of two arcs, each mapping topologically

¹⁰ See an abstract by W. T. Puckett and G. Watson, *Bulletin of the American Mathematical Society*, 43-3-182.

onto \bar{U} . Again these are already edges of the complex on A . If \bar{U} gives Case 2b, the arc \bar{U} is subdivided by the insertion of a vertex at the point $K^1 \cdot K^2$. Denoting the augmented set of vertices in B by G , each component of $B - G$ is such that its inverse under f in A is precisely two open arcs each mapping topologically onto the component in $B - G$. Denote the complex induced on B by G by K_B and the complex induced on A by $f^{-1}(G)$ by K_A , then the transformation f carries edges of K_A into edges of K_B and in topological fashion.

4.4. *Let f be exactly $(2, 1)$ on the connected linear graph A . For each point $x \in B = f(A)$ the relation $o(x) = 1/2 [o(x^1) + o(x^2)]$ holds, where $f^{-1}(x) = x^1 + x^2$.¹¹*

Proof. Suppose A and B have been subdivided into the finite complexes K_A and K_B of the last paragraph. Since f is exactly $(2, 1)$, each simplex in K_B is the topological image of two and only two of the simplexes of K_A . The asserted relation is a direct result of enumeration.

This formula shows that any exactly $(2, 1)$ mapping defined on an arc (or circle) can contain no point of order three, hence, after showing that the arc and circle cannot be exactly $(2, 1)$ images of an arc, we have another proof that it is impossible to define an exactly $(2, 1)$ continuous transformation on an arc [3]. This relation also shows that any exactly $(2, 1)$ image of a simple closed curve is a simple closed curve. By fitting together two simple arcs to form a simple closed curve and defining mappings similar to those of type (α) and (β) , it is easy to show that an exactly $(k, 1)$ mapping on a simple closed curve need not give a simple closed curve for $k > 2$. Since this relation also implies that an exactly $(2, 1)$ transformation cannot be defined on **any** dendritic graph,¹¹ it is of interest to give all of the possibilities (topologically) when A is a simple closed curve.

4.5. *Let f be exactly $(2, 1)$ on the simple closed curve A . Then $B = f(A)$ is a simple closed curve and f is topologically equivalent to either (a) $w = z^2$ on $|z| = 1$, or (b) $w = z^2$ on $|z| = 1$ for $\Re(z) \geq 0$, and $\bar{w} = z^2$ on $|z| = 1$ for $\Re(z) \leq 0$.*

Proof. The transformation $f(A) = B$ is said to be topologically equivalent to $g(A^1) = B^1$ provided there exists a pair of homeomorphisms h and h^1 such that $h(A) = A^1$ and $h^1(B^1) = B$ and $f \equiv h^1gh$ (or $g \equiv (h^1)^{-1}fh^{-1}$) [6].

¹¹ A. D. Wallace pointed out that this implies $2\psi(B) = \psi(A)$, where ψ is the Euler characteristic. Hence if A is a dendritic graph, no exactly $(2, 1)$ transformation can be defined on A . This result has been announced elsewhere. See P. W. Gilbert, abstract 45-11-420, *Bulletin of the American Mathematical Society*.

Two cases are distinguished according as f is interior or not. If f is interior on the simple closed curve A and is $(2, 1)$, and, if it is known further that the image is a simple closed curve, then f is topologically equivalent to $w = z^2$ on $|z| = 1$ [1]. If f is not interior on A , there is a point $x^1 \in A$ and an open set $U \supset x^1$ such that $x = f(x^1)$ is a boundary point of $f(U)$. Let x^2 be the other inverse of x . Let h be a homeomorphism carrying A into the unit circle A^1 in the complex z plane such that $h(x^1) = +1$ and $h(x^2) = -1$. Let h^1 be a homeomorphism carrying the unit circle B^1 of the complex w plane into B such that $h^1(1) = x$. Since f is not interior at x^1 (or x^2), it follows that each of the arcs of A determined by x^1 and x^2 map onto the whole of B under f . Denote by C and D the semi-circles of A^1 , $\Re(z) \geq 0$ and $\Re(z) \leq 0$, respectively. Set $g(z) = (h^1)^{-1}f h^{-1}(z)$. Suppose as C is described from right to left that B^1 is described by $g(z)$ in counterclockwise fashion. Then on $h^{-1}(C)$ the function f is topologically equivalent to $w = z^2$ on $|z| = 1$, $\Re(z) \geq 0$. Since $x = f(x^1)$ is a boundary point of the transform of some open set containing x^1 in A , B^1 must be described in opposite fashion as D is described from left to right. Hence on $h^{-1}(D)$ the function f is topologically equivalent to $\bar{w} = z^2$ on $|z| = 1$, $\Re(z) \leq 0$.

THE UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY

1. G. T. Whyburn, "Interior transformations on compact sets," *Duke Mathematical Journal*, vol. 3 (1937), pp. 370-381.
2. S. Eilenberg, "Sur quelques propriétés des transformations localement homéomorphes," *Fundamenta Mathematicae*, vol. 24 (1935), p. 36.
3. O. G. Harrold, Jr., "The non-existence of a certain type of continuous transformation," *Duke Mathematical Journal*, vol. 5, pp. 789-793.
4. W. T. Puckett, Jr., "Concerning local connectedness under the inverse of certain continuous transformations," *American Journal of Mathematics*, vol. 61 (1939), pp. 750-756. 5. 2.
5. G. T. Whyburn, "Non-alternating transformations," *American Journal of Mathematics*, vol. 56 (1934).
6. G. T. Whyburn, "Completely alternating transformations," *Fundamenta Mathematicae*, vol. 27 (1936), pp. 140-146.

THE CHARACTERIZATION OF PSEUDO-SPHERICAL SETS.*¹

By LEONARD M. BLUMENTHAL and GEORGE R. THURMAN.

1. Introduction. We give in this paper the solution of a fundamental problem in the distance geometry of the n -dimensional sphere (surface) proposed some years ago by Karl Menger.

Defining a semimetric space as a set of abstract elements (*points*), to each pair p, q of which there is attached a non-negative real number pq (*distance*) such that $pq = qp$ and $pq = 0$ if and only if $p = q$, the problem of characterizing *metrically* (i. e., in terms of the distance function) particular semimetric spaces among the whole class of such spaces naturally arises. For some of the more important spaces (e. g., euclidean, hyperbolic, spherical) the existence of a function mapping an arbitrary semimetric space *congruently* (i. e., with preservation of distances) upon the space follows from the congruent embedding in the space of each set of k points of the semimetric space. A space with this property is said to have *congruence order k with respect to semimetric spaces*.² It has been shown that the n -dimensional euclidean, hyperbolic, and spherical spaces have (minimum) congruence order $n + 3$, while, on the other hand, Hilbert space does not have any (finite) congruence order with respect to semimetric spaces.

The problem of determining necessary and sufficient conditions for the congruent embedding of any semimetric space in a given space is thus reducible to a "finite" problem in the case of those spaces possessing a congruence order; for if S has congruence order k with respect to semimetric spaces, then any such space is congruent with a subset of S provided *each set of k points* of the space is congruently embeddable in S . Now the class of semimetric spaces with minimum congruence order $m + 1$ contains a subclass (spaces with *quasi congruence order m*) for each member of which a further reduction in the characterization problem is possible. A space S has quasi congruence order m with respect to semimetric spaces provided any semimetric space *containing more than $m + 1$ points* is congruent with a subset of S whenever each m -tuple

* Received September 30, 1939.

¹ Presented to the Society, December 28, 1938. A brief summary of results appeared in the *Proceedings of the National Academy of Sciences*, vol. 24 (1938), pp. 557-558.

² In this paper the class of comparison spaces is invariably the class of semimetric spaces, and hence the phrase "with respect to semimetric spaces" is frequently omitted.

of its points is congruently embeddable in S . The n -dimensional euclidean and hyperbolic spaces belong to such a subclass of the class of semimetric spaces having minimum congruence order $n + 3$, since each of these spaces has quasi congruence order $n + 2$. On the other hand, the n -dimensional spherical space $S_{n,r}$ (the "surface" of a sphere of radius r in a euclidean space of $n + 1$ dimensions, with geodesic (shorter arc) distance), though it has, as remarked above, minimum congruence order $n + 3$, is not a member of this subclass since the $S_{n,r}$ does not have quasi congruence order $n + 2$.

That this is the case is immediately verified upon noting that the $S_{n,r}$ contains an equilateral set of $n + 2$ points (i. e., a set of $n + 2$ points with all of the $\frac{1}{2}(n + 1)(n + 2)$ mutual distances equal) but does not contain an equilateral $(n + 3)$ -tuple. Hence, if P is a space of arbitrary power exceeding $n + 3$, such that $pq = r \cdot \cos^{-1}(-1/(n + 1))$, (the "side" of an equilateral $(n + 2)$ -tuple of $S_{n,r}$) for $p \neq q$, and $pq = 0$ when $p = q$, ($p, q \in P$), then P is a semimetric space, containing more than $n + 3$ points, which is not congruent with a subset of $S_{n,r}$ though each set of $n + 2$ points of P is congruent with $n + 2$ points of $S_{n,r}$. A semimetric space which is not congruent with a subset of the $S_{n,r}$, though each $n + 2$ of its points may be embedded congruently in the $S_{n,r}$, is called a *pseudo- $S_{n,r}$ set*. As illustrated by the set P defined above, pseudo- $S_{n,r}$ sets may be of arbitrary power exceeding $n + 2$. This is in marked contrast to the analogous pseudo-euclidean sets, for a pseudo- E_n set is restricted to consist of exactly $n + 3$ points, due to the quasi congruence order $n + 2$ property of the E_n . The metric structure of pseudo- E_n sets is readily described.³

The characterization of pseudo- $S_{n,r}$ sets was proposed by Menger in 1931.⁴ The equilateral set P is a pseudo- $S_{n,r}$ set, but are all pseudo- $S_{n,r}$ sets of this simple structure? The principal result of this paper permits us to say that *if a pseudo- $S_{n,r}$ set contains more than $n + 3$ points, and no two of the points have a distance equal to $d = \pi r$, then this query is to be answered essentially in the affirmative*. The meaning of the qualification of the above statement given by the word "essentially" will become clear later.

Pseudo- $S_{n,r}$ sets of exactly $n + 3$ points — it is *proved* (Theorem 6) that no *diametral* pair of points (i. e., two points with distance d) can occur in such a set — have a more varied structure. These sets are described by use of the spherical analogue of the isogonal conjugate transformation of the plane. The case $n = 2$ of the ordinary sphere illustrates all of the essential features.

³ See L. M. Blumenthal, "Distance geometries," *University of Missouri Studies*, vol. 13 (1938), pp. 63-64.

⁴ For $n = 1$ the term pseudo d -cyclic is used. Concerning the characterization of pseudo d -cyclic and pseudo- $S_{2,r}$ sets see "Distance geometries," pp. 74-81.

It is easily seen that a semimetric set of five points p_1, p_2, \dots, p_5 forms a pseudo- $S_{2,r}$ quintuple if and only if the sphere contains five points s_1, s_2, \dots, s_5 such that (1) s_5 is equidistant (with distance $R \neq s_4 s_5$) from the three reflected images $s_4^I, s_4^{II}, s_4^{III}$ of the point s_4 in the great circles $C(s_2, s_3), C(s_1, s_3), C(s_1, s_2)$, respectively, determined by the *independent* (i. e., not on a great circle) points s_1, s_2, s_3 and (2) the mutual distances of the points p_1, p_2, \dots, p_5 equal the corresponding distances of the points s_1, s_2, \dots, s_5 *except for the distance* $p_4 p_5$, which equals R instead of $s_4 s_5$.⁵ Thus, the four distances $p_1 p_5, p_2 p_5, p_3 p_5$, and $p_4 p_5 = R$ are functions of the six mutual distances of the four points p_1, p_2, p_3, p_4 . The actual expressions for these four distances have not been computed (nor has the analogous computation for pseudo-plane sets been made). One should note here a complication that arises in the spherical analogue of the isogonal conjugate transformation which is not present in the plane. A point s_5 of the $s_{2,r}$ is not uniquely determined by being equidistant from the three points $s_4^I, s_4^{II}, s_4^{III}$ —a pair of diametral points satisfies this condition. Thus, the transformation on the sphere is not one-to-one, as it is (apart from certain exceptional points) in the plane, but is one-to-two, or rather, two-to-two, since a diametral pair is transformed into a diametral pair.

The process sketched for pseudo- $S_{2,r}$ quintuples is representative of the procedure for pseudo- $S_{n,r}$ ($n+3$)-tuples. The determination of their metric structure is, then, so closely related to the analogous problem for pseudo- E_n ($n+3$)-tuples as to present nothing essentially new. On the other hand, one may surely expect quite different results when pseudo- $S_{n,r}$ sets of more than $n+3$ points are considered, for their euclidean analogues (pseudo- E_n sets of more than $n+3$ points) do not exist. Furthermore, the complication due to the one-to-two character of the transformation described above makes itself felt only for pseudo- $S_{n,r}$ sets of more than $n+3$ points. Such sets may contain diametral points unless the contrary is explicitly assumed.

2. Basic and derived properties of the $S_{n,r}$ and its subspaces. Many properties of the $S_{n,r}$ and its k -dimensional subspaces $S_{k,r}$, $0 \leq k \leq n$, (the sections of the $S_{n,r}$ made by $(k+1)$ -dimensional hyperplanes through the center of the sphere in E_{n+1} whose "surface" is the $S_{n,r}$) are needed for the development of this paper. Looking towards the "abstraction" of our problem made in Section 4, we isolate here the *five basic properties of the $S_{n,r}$ that*

⁵ Throughout this paper the points of pseudo- $S_{n,r}$ sets are denoted by the letters p and q , while the points of $S_{n,r}$ are symbolized by the letters s and t .

suffice to demonstrate all the additional properties of the $S_{n,r}$ that we need for this investigation.⁶

I. The determinant

$$\Delta_{n+2}(s_1, s_2, \dots, s_{n+2}) = |\cos(s_i s_j / r)|, \\ (i, j = 1, 2, \dots, n+2),$$

vanishes for each set of $n+2$ points s_1, s_2, \dots, s_{n+2} of $S_{n,r}$.

II. There exists at least one set of $n+1$ points of $S_{n,r}$ whose determinant Δ_{n+1} does not vanish.

III. Each finite subset of $S_{n,r}$ has a non-negative determinant Δ .

Remark. It is easy to show that the dependence (independence) of a finite set s_1, s_2, \dots, s_k of k points of $S_{n,r}$ is equivalent to the vanishing (non-vanishing) of the determinant Δ_k of the k points. We call k points of a semimetric space independent (dependent) if they are congruent with k independent (dependent) points of $S_{n,r}$.

IV. If $s_1, s_2, \dots, s_{k+1} \approx t_1, t_2, \dots, t_{k+1}$ are two congruent sets of $k+1$ points (not necessarily pairwise distinct) of two (coincident or distinct) k -dimensional subspaces of $S_{n,r}$, $k \leq n$, then to each point s of the subspace containing s_1, s_2, \dots, s_{k+1} there corresponds at least one point t of the subspace containing t_1, t_2, \dots, t_{k+1} such that

$$s_1, s_2, \dots, s_{k+1}, s \approx t_1, t_2, \dots, t_{k+1}, t.$$

V. If s_1, s_2, \dots, s_k are k independent points of $S_{k,r}$, ($k=1, 2, \dots, n$), then corresponding to each point s of $S_{k,r}$ independent of s_1, s_2, \dots, s_k (i. e., $\Delta_{k+1}(s_1, s_2, \dots, s_k, s)$ does not vanish), there is at least one point s' of $S_{k,r}$ such that $s' \neq s$ and $s_1, s_2, \dots, s_k, s \approx s_1, s_2, \dots, s_k, s'$.

We list now for convenience the *derived properties* of the $S_{n,r}$ to which reference will be made in the next section:

(a). A semimetric set of $k+2$ points is congruent with $k+2$ points of $S_{k,r}$, $k \leq n$, if and only if each $k+1$ of the points are congruent with $k+1$ points of the $S_{k,r}$ and the determinant Δ_{k+2} of the $k+2$ points vanishes.

⁶ The derivation of these additional properties from the five basic ones follows closely the methods of an earlier paper (L. M. Blumenthal, "The geometry of a class of semimetric spaces," *Tôhoku Mathematical Journal*, vol. 43 (1937), pp. 205-224).

⁷ This notation signifies that $s_i s_j = t_i t_j$, ($i, j = 1, 2, \dots, k+1$).

(b). A semimetric set of $k+3$ points is congruent with $k+3$ points of $S_{k,r}$, $k \leq n$, if and only if each $k+2$ of the points are congruent with $k+2$ points of the $S_{k,r}$ and the determinant Δ_{k+3} of the $k+3$ points vanishes.⁸

(c). If $s_1, s_2, \dots, s_{k+1} \approx t_1, t_2, \dots, t_{k+1}$ are two congruent sets of $k+1$ independent points of two (coincident or distinct) k -dimensional subspaces, $k \leq n$, and if s, s' are points of the subspace containing s_1, s_2, \dots, s_{k+1} , while t, t' are points of the subspace containing t_1, t_2, \dots, t_{k+1} such that

$$\begin{aligned}s_1, s_2, \dots, s_{k+1}, s &\approx t_1, t_2, \dots, t_{k+1}, t, \\ s_1, s_2, \dots, s_{k+1}, s' &\approx t_1, t_2, \dots, t_{k+1}, t',\end{aligned}$$

then $ss' = tt'$.

Remark 1. If $s_1, s_2, \dots, s_{k+1} \approx t_1, t_2, \dots, t_{k+1}$ are two congruent sets of $k+1$ independent points of two (coincident or distinct) k -dimensional subspaces, $k \leq n$, then to each point s of the subspace containing the first set of $k+1$ points there corresponds exactly one point t of the subspace containing the second set of $k+1$ points such that

$$s_1, s_2, \dots, s_{k+1}, s \approx t_1, t_2, \dots, t_{k+1}, t.$$

Remark 2. There is at most one point of $S_{k,r}$ with prescribed distances from $k+1$ independent points of $S_{k,r}$, $k \leq n$.

(d). Any subset of an independent set of points is an independent set of points.

(e). If s_1, s_2, \dots, s_k are k independent points of $S_{k,r}$, ($k=1, 2, \dots, n$), then corresponding to each point s of $S_{k,r}$ which is independent of them there is exactly one point s' of $S_{k,r}$ such that $s' \neq s$ and

$$s_1, s_2, \dots, s_k, s \approx s_1, s_2, \dots, s_k, s'.^9$$

(f). Let s_1, s_2, \dots, s_k be k independent points of $S_{k,r}$, ($k=1, 2, \dots, n$), and let s, s' and t, t' be two pairs of points of $S_{k,r}$ such that t is either dependent on s_1, s_2, \dots, s_k or distinct from s , and t' is either dependent on s_1, s_2, \dots, s_k or distinct from s' , while

$$\begin{aligned}s_1, s_2, \dots, s_k, s &\approx s_1, s_2, \dots, s_k, t, \\ s_1, s_2, \dots, s_k, s' &\approx s_1, s_2, \dots, s_k, t'.\end{aligned}$$

Then $ss' = tt'$.

⁸ *Properties (a), (b)* are well-known theorems in the distance geometry of the $S_{n,r}$ (see "Distance geometries," p. 73). That they may be obtained by using merely *Properties I-V* has not been recorded heretofore.

⁹ The point s has a single image $s' \neq s$ when "reflected" in the $(k-1)$ -dimensional subspace determined by the k independent points s_1, s_2, \dots, s_k .

(g). Let s_1, s_2, \dots, s_{k+1} be $k+1$ independent points of $S_{k,r}$, $k \leq n$, and let s be a point of $S_{k,r}$ not common to any two of the $k+1$ subspaces $S_{k-1,r}$ determined by the k points

$$s_1, s_2, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+1}, \quad (i = 1, 2, \dots, k+1).$$

Denote by $s^{(i)}$ a point of $S_{k,r}$ such that

$$s_1, s_2, \dots, s_{i-1}, s^{(i)}, s_{i+1}, \dots, s_{k+1} \approx s_1, s_2, \dots, s_{i-1}, s, s_{i+1}, \dots, s_{k+1}, \quad (i = 1, 2, \dots, k+1),$$

where $s^{(i)} \neq s$ if s is independent of $s_1, s_2, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+1}$. Then there are not two independent points s', t' of $S_{k,r}$ such that

$$s's^{(1)} = s's^{(2)} = \dots = s's^{(k+1)} \quad \text{and} \quad t's^{(1)} = t's^{(2)} = \dots = t's^{(k+1)}.$$

Remark. There are at most two points of $S_{k,r}$ satisfying the conditions of property (g).

(h). The $S_{k,r}$, $k \leq n$, has minimum congruence order $k+3$.¹⁰

(i). If s_1, s_2, \dots, s_{k+1} are $k+1$ independent points of $S_{k,r}$, $k \leq n$, and s is a point of $S_{k,r}$ such that for each integer i , ($i = 1, 2, \dots, k$) the points $s_1, s_2, \dots, s_{i-1}, s, s_{i+1}, \dots, s_k, s_{k+1}$ are dependent, then the points s, s_{k+1} are dependent.

(j). If s, t, u are any three distinct points of $S_{k,r}$, $k \leq n$, such that $st = tu = su$, and if s_1, s_2, \dots, s_k are k points of $S_{k,r}$ such that

$$s_i s = s_i t = s_i u, \quad (i = 1, 2, \dots, k),$$

then the points s_1, s_2, \dots, s_k are dependent.

(k). Let s, t be any two distinct points of $S_{k,r}$, ($k = 1, 2, \dots, n$), and let s_1, s_2, \dots, s_k be k independent points of $S_{k,r}$ such that $s_i s = s_i t$, ($i = 1, 2, \dots, k$). The $(k-1)$ -dimensional subspace $S_{k-1,r}$ determined by s_1, s_2, \dots, s_k is the locus of points of $S_{k,r}$ equidistant from s and t .

(l). Two distinct k -dimensional subspaces $S_{k,r}$, $k \leq n$, can have at most k independent points in common.

3. The characterization of pseudo- $S_{k,r}$ sets, $k \leq n$. We deduce first some preliminary theorems concerning pseudo- $S_{k,r}$ sets, $k \leq n$, that point the way to the desired characterization theorems.

¹⁰ "Distance geometries," p. 73.

THEOREM 1. *A pseudo- $S_{k,r}$ $(k+3)$ -tuple p_1, p_2, \dots, p_{k+3} contains at least one independent set of $k+1$ points.¹¹*

Proof. Since p_1, p_2, \dots, p_{k+3} is a pseudo- $S_{k,r}$ set, each $(k+2)$ -tuple contained in these points is congruent with $k+2$ points of $S_{k,r}$. It follows (*property (a)*) that the determinant Δ_{k+2} of each $(k+2)$ -tuple vanishes. Suppose, now, that each set of $k+1$ of the $k+3$ points is a dependent set. Then the determinant $\Delta_{k+3}(p_1, p_2, \dots, p_{k+3})$ has all principal minors of orders $k+1$ and $k+2$ equal to zero, and consequently vanishes. It follows (*property (b)*) that the $k+3$ points are congruent with a subset of the $S_{k,r}$, which contradicts the hypothesis that they form a pseudo- $S_{k,r}$ set, and establishes the theorem.

THEOREM 2. *The determinant $\Delta_{k+3}(p_1, p_2, \dots, p_{k+3})$ of a pseudo- $S_{k,r}$ $(k+3)$ -tuple is negative.*

Proof. As seen in the proof of Theorem 1, the determinant

$$\Delta_{k+3}(p_1, p_2, \dots, p_{k+3})$$

does not vanish. Let the points be so labelled that p_1, p_2, \dots, p_{k+1} is an independent $(k+1)$ -tuple. Since

$$\Delta_{k+2}(p_1, p_2, \dots, p_{k+2}) = 0,$$

we have¹²

$$\Delta_{k+3}(p_1, p_2, \dots, p_{k+3}) = \frac{-[k+2, k+3]^2}{\Delta_{k+1}(p_1, p_2, \dots, p_{k+1})},$$

where $[k+2, k+3]$ denotes the co-factor of the element in the $(k+2)$ -nd row and $(k+3)$ -rd column of $\Delta_{k+3}(p_1, p_2, \dots, p_{k+3})$. The points p_1, p_2, \dots, p_{k+1} being independent and congruent to $k+1$ points of $S_{k,r}$ implies $\Delta_{k+1}(p_1, p_2, \dots, p_{k+1}) > 0$, (*property III*), and the theorem is proved.

THEOREM 3. *If p_1, p_2, \dots, p_{k+3} form a pseudo- $S_{k,r}$ set, with the points*

¹¹ We shall suppose the index k to assume the values $k=0, 1, 2, \dots, n$ except when it is stated otherwise.

¹² The expansion of a $(k+3)$ -rd order symmetric determinant

$$|a_{ij}| = ([k+2, k+2] \cdot [k+3, k+3] - [k+2, k+3]^2) / \begin{vmatrix} k+2, k+2 & k+2, k+3 \\ k+3, k+2 & k+3, k+3 \end{vmatrix}$$

(a special case of Jacobi's theorem), is particularly useful whenever, as frequently happens in this paper, one of the co-factors $[k+2, k+2]$, $[k+3, k+3]$ vanishes.

p_1, p_2, \dots, p_{k+1} independent, then the $S_{k,r}$ contains $k+3$ points s_1, s_2, \dots, s_{k+3} such that

$$p_1, p_2, \dots, p_{k+1}, p_{k+2} \approx s_1, s_2, \dots, s_{k+1}, s_{k+2},$$

$$p_1, p_2, \dots, p_{k+1}, p_{k+3} \approx s_1, s_2, \dots, s_{k+1}, s_{k+3},$$

and $p_{k+2}p_{k+3} \neq s_{k+2}s_{k+3}$.

Proof. Since p_1, p_2, \dots, p_{k+3} form a pseudo- $S_{k,r}$ set, there exist two sets $s_1, s_2, \dots, s_{k+1}, s_{k+2}$ and $t_1, t_2, \dots, t_{k+1}, t_{k+3}$ of $k+2$ points of $S_{k,r}$ such that

$$p_1, p_2, \dots, p_{k+1}, p_{k+2} \approx s_1, s_2, \dots, s_{k+1}, s_{k+2},$$

$$p_1, p_2, \dots, p_{k+1}, p_{k+3} \approx t_1, t_2, \dots, t_{k+1}, t_{k+3}.$$

The points p_1, p_2, \dots, p_{k+1} being independent, then $\{s_i\}$ and $\{t_i\}$, ($i=1, 2, \dots, k+1$), are two congruent sets of $k+1$ independent points of $S_{k,r}$, and hence (*Remark 1, property (c)*) the $S_{k,r}$ contains exactly one point s_{k+3} such that

$$t_1, t_2, \dots, t_{k+1}, t_{k+3} \approx s_1, s_2, \dots, s_{k+1}, s_{k+3}.$$

Then we have

$$p_1, p_2, \dots, p_{k+1}, p_{k+3} \approx s_1, s_2, \dots, s_{k+1}, s_{k+3},$$

and since the $k+3$ points p_1, p_2, \dots, p_{k+3} form a pseudo- $S_{k,r}$ set, the distance $p_{k+2}p_{k+3}$ does not equal the distance $s_{k+2}s_{k+3}$.

The $k+3$ points s_1, s_2, \dots, s_{k+3} of $S_{k,r}$ are said to be "almost congruent" to the pseudo- $S_{k,r}$ set p_1, p_2, \dots, p_{k+3} .

THEOREM 4. Let p_1, p_2, \dots, p_{k+3} form a pseudo- $S_{k,r}$ $(k+3)$ -tuple with the independent $(k+1)$ -tuple p_1, p_2, \dots, p_{k+1} . Then the $k+1$ points p_2, p_3, \dots, p_{k+2} are independent.

Proof. Let s_1, s_2, \dots, s_{k+3} be $k+3$ points of $S_{k,r}$ almost congruent to p_1, p_2, \dots, p_{k+3} ; i. e.,

$$p_1, p_2, \dots, p_{k+1}, p_{k+2} \approx s_1, s_2, \dots, s_{k+1}, s_{k+2},$$

$$p_1, p_2, \dots, p_{k+1}, p_{k+3} \approx s_1, s_2, \dots, s_{k+1}, s_{k+3},$$

and $p_{k+2}p_{k+3} \neq s_{k+2}s_{k+3}$. Since each $k+2$ points of the set p_1, p_2, \dots, p_{k+3} are congruent with $k+2$ points of $S_{k,r}$, we have

$$p_2, p_3, \dots, p_{k+2}, p_{k+3} \approx t_2, t_3, \dots, t_{k+2}, t_{k+3},$$

and hence

$$t_2, t_3, \dots, t_{k+1}, t_{k+2} \approx s_2, s_3, \dots, s_{k+1}, s_{k+2},$$

with the points s_2, s_3, \dots, s_{k+1} being independent since they belong to the independent $(k+1)$ -tuple s_1, s_2, \dots, s_{k+1} (*property (d)*).

Property IV applied to the congruence

$$t_2, t_3, \dots, t_{k+1} \approx s_2, s_3, \dots, s_{k+1}$$

entails the existence of two points s, s' of $S_{k,r}$ such that

$$t_2, t_3, \dots, t_{k+1}, t_{k+2}, t_{k+3} \approx s_2, s_3, \dots, s_{k+1}, s, s'.$$

Then

$$s_2, s_3, \dots, s_{k+1}, s \approx t_2, t_3, \dots, t_{k+1}, t_{k+2} \approx s_2, s_3, \dots, s_{k+1}, s_{k+2}.$$

Suppose, now, that the points $p_2, p_3, \dots, p_{k+1}, p_{k+2}$ are dependent. Then $s_2, s_3, \dots, s_{k+1}, s_{k+2}$ are dependent, and applying *Remark 2, property (c)* to the $S_{k-1,r}$ determined by the k independent points s_2, s_3, \dots, s_{k+1} , the above congruence gives $s = s_{k+2}$, and hence $t_2, t_3, \dots, t_{k+1}, t_{k+2}, t_{k+3}$ are congruent with $s_2, s_3, \dots, s_{k+1}, s_{k+2}, s'$ in the usual order.

Now

$$s_2, s_3, \dots, s_{k+1}, s_{k+3} \approx p_2, p_3, \dots, p_{k+1}, p_{k+3} \approx t_2, t_3, \dots, t_{k+1}, t_{k+3},$$

which gives $s_2, s_3, \dots, s_{k+1}, s_{k+3} \approx s_2, s_3, \dots, s_{k+1}, s'$. If s_{k+3} is not dependent on s_2, s_3, \dots, s_{k+1} , then by *property (e)* the point s' may be distinct from s_{k+3} . But then the last congruence together with the congruence

$$s_2, s_3, \dots, s_{k+1}, s_{k+2} \approx s_2, s_3, \dots, s_{k+1}, s_{k+2}$$

shows that $s_{k+2}s_{k+3} = s_{k+2}s'$, according to *property (f)*. If, on the other hand, s_{k+3} is dependent on s_2, s_3, \dots, s_{k+1} , then the congruence

$$s_2, s_3, \dots, s_{k+1}, s_{k+3} \approx s_2, s_3, \dots, s_{k+1}, s'$$

evidently implies $s' = s_{k+3}$, and $s_{k+2}s_{k+3} = s_{k+2}s'$ as before. Hence, in any case,

$$s_{k+2}s_{k+3} = s_{k+2}s' = t_{k+2}t_{k+3} = p_{k+2}p_{k+3},$$

which gives the desired contradiction and establishes the theorem.

Remark. It is clear that the method of the preceding theorem may be used to show that each of the $(k+1)$ -tuples

$$p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_{k+1}, p_{k+2}; \quad p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_{k+1}, p_{k+3}, \\ (i = 1, 2, \dots, k+1),$$

is independent.

THEOREM 5. Let $p_1, p_2, \dots, p_{k+2}, p_{k+3}$ form a pseudo- $S_{k,r}$ set with the independent $(k+1)$ -tuples $p_1, p_2, \dots, p_k, p_{k+1}$;

$$p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_{k+1}, p_{k+2}; p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_{k+1}, p_{k+3}, \\ (i = 1, 2, \dots, k+1).$$

Then the $(k+1)$ -tuple $p_3, p_4, \dots, p_{k+2}, p_{k+3}$ is independent.

Proof. The proof of this theorem follows the lines of the proof of the preceding theorem, with the independent $(k+1)$ -tuple $p_1, p_3, p_4, \dots, p_{k+1}, p_{k+2}$ in the rôle of the $(k+1)$ -tuple p_1, p_2, \dots, p_{k+1} . Thus, the $S_{k,r}$ contains $k+3$ points $s_1, s_2, \dots, s_{k+2}, s_{k+3}$ such that

$$p_1, p_3, p_4, \dots, p_{k+1}, p_{k+2}, p_{k+3} \approx s_1, s_3, s_4, \dots, s_{k+1}, s_{k+2}, s_{k+3}, \\ p_1, p_3, p_4, \dots, p_{k+1}, p_{k+2}, p_2 \approx s_1, s_3, s_4, \dots, s_{k+1}, s_{k+2}, s_2,$$

with $p_2 p_{k+3} \neq s_2 s_{k+3}$, and the same procedure as in Theorem 4 leads to the desired result.

Remark. In a similar manner, it is seen that the $(k+1)$ -tuples

$$p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_{j-1}, p_{j+1}, \dots, p_{k+1}, p_{k+2}, p_{k+3}, \\ (i, j = 1, 2, \dots, k+1; i \neq j),$$

are all independent $(k+1)$ -tuples.

We have thus proved the following useful theorem:

THEOREM 6. If $p_1, p_2, \dots, p_{k+2}, p_{k+3}$ form a pseudo- $S_{k,r}$ set then each of the $\frac{1}{2}(k+2)(k+3)$ sets of $k+1$ points contained in this $(k+3)$ -tuple is an independent set.

Let p_1, p_2, \dots, p_{k+3} form a pseudo- $S_{k,r}$ $(k+3)$ -tuple, and let s_1, s_2, \dots, s_{k+3} be $k+3$ points of $S_{k,r}$ almost congruent to them. Consider, now, the points

$$p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_{k+3}, \quad (i = 1, 2, \dots, k+1).$$

The $S_{k,r}$ contains $k+2$ points $t_1, t_2, \dots, t_{i-1}, t_{i+1}, \dots, t_{k+3}$ such that for each $i = 1, 2, \dots, k+1$,

$$p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_{k+3} \approx t_1, t_2, \dots, t_{i-1}, t_{i+1}, \dots, t_{k+3}.$$

Then

$$s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+1}, s_{k+2}$$

is congruent with the points

$$t_1, t_2, \dots, t_{i-1}, t_{i+1}, \dots, t_{k+1}, t_{k+2},$$

with each of the sets of $k+1$ points

$$s_1, s_2, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+1}, s_{k+2}, \quad (i = 1, 2, \dots, k+1),$$

being independent (Theorem 6). Hence, by Remark 1, property (c), the $S_{k,r}$ contains exactly one point $s_{k+3}^{(4)}$ such that

$$s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+2}, s_{k+3}^{(4)} \approx t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_{k+2}, t_{k+3}, \\ (i = 1, 2, \dots, k+1),$$

and hence

$$p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{k+2}, p_{k+3} \approx s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+2}, s_{k+3}^{(i)}, \\ (i = 1, 2, \dots, k+1).$$

Since

$$s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+1}, s_{k+3}^{(i)} \approx p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{k+1}, p_{k+3}, \\ \approx s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+1}, s_{k+3},$$

the point $s_{k+3}^{(4)}$ has the same distances from the points

$$s_1, s_2, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+1}$$

as the point s_{k+3} , but $s_{k+3}^{(4)} \neq s_{k+3}$ for

$$s_{k+2}s_{k+3}^{(4)} = p_{k+2}p_{k+3} \neq s_{k+2}s_{k+3}.$$

We have

$$s_{k+2}s_{k+3}^{(1)} = s_{k+2}s_{k+3}^{(2)} = \dots = s_{k+2}s_{k+3}^{(k+1)},$$

for each of these distances equals $p_{k+2}p_{k+3}$. Now a point s_{k+2} is not determined uniquely by these inequalities, but by property (g) and its accompanying Remark there is at most one other point s_{k+2}^* equidistant from the $k+1$ points $s_{k+3}^{(1)}, s_{k+3}^{(2)}, \dots, s_{k+3}^{(k+1)}$ and $s_{k+2}s_{k+2}^* = d$.

In a similar manner it is seen that the $S_{k,r}$ contains $k+1$ points $s_{k+2}^{(1)}, s_{k+2}^{(2)}, \dots, s_{k+2}^{(k+1)}$ such that

$$p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{k+1}, p_{k+2} \approx s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+1}, s_{k+2}^{(i)}, \\ (i = 1, 2, \dots, k+1),$$

with each of the distances $s_{k+2}^{(i)}s_{k+3}$, ($i = 1, 2, \dots, k+1$), equal to $p_{k+2}p_{k+3}$. Thus s_{k+3} is equidistant from the $k+1$ points $s_{k+2}^{(1)}, s_{k+2}^{(2)}, \dots, s_{k+2}^{(k+1)}$, and there are at most two points (diametral) satisfying this requirement.

THEOREM 7. Let $P \equiv (p_1, p_2, \dots, p_{k+3})$ and $Q \equiv (q_1, q_2, \dots, q_{k+3})$ be two pseudo- $S_{k,r}$ ($k+3$)-tuples such that

$$p_1, p_2, \dots, p_{k+2} \approx q_1, q_2, \dots, q_{k+2}.$$

Then either $p_i p_{k+3} = q_i q_{k+3}$, ($i = 1, 2, \dots, k+2$), and the two ($k+3$)-tuples are congruent, or

$$\cos(p_i p_{k+3}/r) + \cos(q_i q_{k+3}/r) = 0, \quad (i = 1, 2, \dots, k+2).^{13}$$

Proof. From the hypothesis of the theorem we may write the following congruences:

$$p_1, p_2, \dots, p_{k+1}, p_{k+2} \approx s_1, s_2, \dots, s_{k+1}, s_{k+2},$$

$$p_1, p_2, \dots, p_{k+1}, p_{k+3} \approx s_1, s_2, \dots, s_{k+1}, s_{k+3},$$

$$q_1, q_2, \dots, q_{k+1}, q_{k+2} \approx s_1, s_2, \dots, s_{k+1}, s_{k+2},$$

$$q_1, q_2, \dots, q_{k+1}, q_{k+3} \approx s_1, s_2, \dots, s_{k+1}, t_{k+3},$$

where the points on the right-hand side of these congruences are in $S_{k,r}$ and

$$p_{k+2} p_{k+3} \neq s_{k+2} s_{k+3}, \quad q_{k+2} q_{k+3} \neq s_{k+2} t_{k+3}.$$

From the first two congruences follow, as has been seen, the existence of points $s_{k+2}^{(i)}$, ($i = 1, 2, \dots, k+1$), of $S_{k,r}$ such that

$$p_{k+2} p_{k+3} = s_{k+2}^{(1)} s_{k+3} = s_{k+2}^{(2)} s_{k+3} = \dots = s_{k+2}^{(k+1)} s_{k+3}.$$

Similarly, from the last two congruences, we may write

$$q_{k+2} q_{k+3} = s_{k+2}^{(1)} t_{k+3} = s_{k+2}^{(2)} t_{k+3} = \dots = s_{k+2}^{(k+1)} t_{k+3}.$$

It follows (*property (g)*) that either $s_{k+3} = t_{k+3}$ or $s_{k+3} t_{k+3} = d$. In the first case, clearly $p_i p_{k+3} = q_i q_{k+3}$, ($i = 1, 2, \dots, k+2$), and $P \approx Q$. In the second case, s_{k+3} and t_{k+3} being diametral implies (*property III*) that the determinant $\Delta_3(s_i, s_{k+3}, t_{k+3}) = 0$, and hence (upon expanding)

$$\cos(s_i s_{k+3}/r) + \cos(s_i t_{k+3}/r) = 0, \quad (i = 1, 2, \dots, k+1).$$

Then, by the above congruences,

$$\cos(p_i p_{k+3}/r) + \cos(q_i q_{k+3}/r) = 0, \quad (i = 1, 2, \dots, k+1).$$

Finally, $s_{k+3} t_{k+3} = d$ implies that $\Delta_3(s_{k+2}^{(1)}, s_{k+3}, t_{k+3})$ vanishes; i. e.,

$$\cos(s_{k+2}^{(1)} s_{k+3}/r) + \cos(s_{k+2}^{(1)} t_{k+3}/r) = 0.$$

Then

$$\cos(p_{k+2} p_{k+3}/r) + \cos(q_{k+2} q_{k+3}/r) = 0,$$

and the theorem is proved.

LEMMA. If a pseudo- $S_{k,r}$ set P of $k+4$ points p_1, p_2, \dots, p_{k+4} , has no pair of its points diametral, then the set contains at least three pseudo- $S_{k,r}$ ($k+3$)-tuples.

¹³ With a view to the "abstraction" of the problem treated in Section 4, we write $\cos(p_i p_{k+3}/r) + \cos(q_i q_{k+3}/r) = 0$ rather than $p_i p_{k+3} + q_i q_{k+3} = d$.

Proof. Since, by *property (h)*, the $S_{k,r}$ has congruence order $k+3$, the set P contains at least one pseudo- $S_{k,r}$ $(k+3)$ -tuple. The labelling may be assumed so that p_1, p_2, \dots, p_{k+3} is a pseudo- $S_{k,r}$ set. In case P does not contain at least two $(k+3)$ -tuples congruent with $k+3$ points of $S_{k,r}$, the lemma is surely valid. In the contrary case, let $p_1, p_2, \dots, p_{k+1}, p_{k+2}, p_{k+4}$ and $p_1, p_2, \dots, p_{k+1}, p_{k+3}, p_{k+4}$ be congruent with two $(k+3)$ -tuples of $S_{k,r}$. Since p_1, p_2, \dots, p_{k+3} is a pseudo- $S_{k,r}$ set, we have

$$\begin{aligned} p_1, p_2, \dots, p_{k+1}, p_{k+2} &\approx s_1, s_2, \dots, s_{k+1}, s_{k+2}, \\ p_1, p_2, \dots, p_{k+1}, p_{k+3} &\approx s_1, s_2, \dots, s_{k+1}, s_{k+3}, \end{aligned}$$

with $p_{k+2}p_{k+3} \neq s_{k+2}s_{k+3}$, and it follows that

$$(I) \quad \begin{aligned} p_1, p_2, \dots, p_{k+1}, p_{k+2}, p_{k+4} &\approx s_1, s_2, \dots, s_{k+1}, s_{k+2}, s_{k+4}, \\ p_1, p_2, \dots, p_{k+1}, p_{k+3}, p_{k+4} &\approx s_1, s_2, \dots, s_{k+1}, s_{k+3}, s_{k+4}, \end{aligned}$$

and the point s_{k+4} of $S_{k,r}$ is *uniquely determined* since its distances from the $k+1$ independent points s_1, s_2, \dots, s_{k+1} of $S_{k,r}$ are fixed. Now, by hypothesis, each pair of points of P is independent (i.e., no two points of P are coincident or diametral) and hence the points s_{k+4} and s_i , ($i=1, 2, \dots, k+3$), are independent. It follows that at least two of the $k+1$ sets

$$s_1, s_2, \dots, s_{i-1}, s_{i+1}, \dots, s_{k+1}, s_{k+4}, \quad (i=1, 2, \dots, k+1),$$

are independent $(k+1)$ -tuples, for in the contrary case, the k dependent $(k+1)$ -tuples have, in addition to the point s_{k+4} one of the points s_1, s_2, \dots, s_{k+1} in common. This point is then, by *property (i)*, diametral to (or coincident with) the point s_{k+4} , in contradiction to the preceding remark.

Let, then, $s_2, s_3, \dots, s_{k+1}, s_{k+4}$ and $s_1, s_3, \dots, s_{k+1}, s_{k+4}$ be independent $(k+1)$ -tuples. We show that the two $(k+3)$ -tuples

$$p_2, p_3, \dots, p_{k+1}, p_{k+2}, p_{k+3}, p_{k+4}$$

and

$$p_1, p_3, \dots, p_{k+1}, p_{k+2}, p_{k+3}, p_{k+4}$$

are pseudo- $S_{k,r}$ sets. (A similar procedure gives the desired result in case two other $(k+1)$ -tuples are independent.)

We make the assumption that $p_2, p_3, \dots, p_{k+1}, p_{k+2}, p_{k+3}, p_{k+4}$ are congruent with $k+3$ points of $S_{k,r}$ and show that this assumption leads to a contradiction. (The other $(k+3)$ -tuple is treated in the same manner.)

Suppose, then, that

$$p_2, p_3, \dots, p_{k+1}, p_{k+2}, p_{k+3}, p_{k+4} \approx t_2, t_3, \dots, t_{k+1}, t_{k+2}, t_{k+3}, t_{k+4},$$

of $S_{k,r}$. Then (using congruences (I)) it follows that

$$s_2, s_3, \dots, s_{k+1}, s_{k+2}, s_{k+4} \approx t_2, t_3, \dots, t_{k+1}, t_{k+2}, t_{k+4},$$

$$s_2, s_3, \dots, s_{k+1}, s_{k+3}, s_{k+4} \approx t_2, t_3, \dots, t_{k+1}, t_{k+3}, t_{k+4},$$

with $s_2, s_3, \dots, s_{k+1}, s_{k+4}$ an independent $(k+1)$ -tuple. It follows that $s_{k+2}s_{k+3} = t_{k+2}t_{k+3} = p_{k+2}p_{k+3}$ (the first equality resulting from property (c)) gives the desired contradiction, and proves that the $(k+3)$ -tuple

$$p_2, p_3, \dots, p_{k+1}, p_{k+2}, p_{k+3}, p_{k+4},$$

is a pseudo- $S_{k,r}$ set.

THEOREM 8. *If a pseudo- $S_{k,r}$ $(k+4)$ -tuple P has no pair of its points diametral, then each $(k+3)$ -tuple contained in P is a pseudo- $S_{k,r}$ set.*

Proof. From the Lemma, P contains at least three pseudo- $S_{k,r}$ $(k+3)$ -tuples, say

$$p_1, p_2, \dots, p_{k+1}, p_{k+2}, p_{k+3}; \quad p_1, p_2, \dots, p_{k+1}, p_{k+2}, p_{k+4};$$

$$p_1, p_2, \dots, p_{k+1}, p_{k+3}, p_{k+4}.$$

Applying Theorem 7 to the first and second of these pseudo- $S_{k,r}$ sets, and then to the first and third of these sets, we see that the distances determined by the points of P satisfy one of the following four sets of relations:

$$\begin{aligned} \text{Case I.} \quad & p_{k+2}p_{k+3} = p_{k+3}p_{k+4} = p_{k+2}p_{k+4}, \\ & p_i p_{k+2} = p_i p_{k+3} = p_i p_{k+4}, \quad (i = 1, 2, \dots, k+1). \end{aligned}$$

$$\begin{aligned} \text{Case II.} \quad & \cos(p_{k+2}p_{k+3}/r) = -\cos(p_{k+3}p_{k+4}/r) = -\cos(p_{k+2}p_{k+4}/r), \\ & \cos(p_i p_{k+2}/r) = \cos(p_i p_{k+3}/r) = -\cos(p_i p_{k+4}/r), \\ & (i = 1, 2, \dots, k+1). \end{aligned}$$

$$\begin{aligned} \text{Case III.} \quad & \cos(p_{k+2}p_{k+3}/r) = -\cos(p_{k+3}p_{k+4}/r) = \cos(p_{k+2}p_{k+4}/r), \\ & \cos(p_i p_{k+2}/r) = -\cos(p_i p_{k+3}/r) = -\cos(p_i p_{k+4}/r), \\ & (i = 1, 2, \dots, k+1). \end{aligned}$$

$$\begin{aligned} \text{Case IV.} \quad & \cos(p_{k+2}p_{k+3}/r) = \cos(p_{k+3}p_{k+4}/r) = -\cos(p_{k+2}p_{k+4}/r), \\ & \cos(p_i p_{k+2}/r) = -\cos(p_i p_{k+3}/r) = \cos(p_i p_{k+4}/r), \\ & (i = 1, 2, \dots, k+1). \end{aligned}$$

To show, for example, that $p_2, p_3, \dots, p_{k+1}, p_{k+2}, p_{k+3}, p_{k+4}$ is a pseudo- $S_{k,r}$ $(k+3)$ -tuple, assume the contrary and let $s_2, s_3, \dots, s_{k+1}, s_{k+2}, s_{k+3}, s_{k+4}$ be $k+3$ points of $S_{k,r}$ congruent to them. Then the distances determined by these $k+3$ points of the $S_{k,r}$ satisfy one of the above four sets of relations (with the index i taking on the values $2, 3, \dots, k+1$).

Case I. Applying *property (j)* to s_2, s_3, \dots, s_{k+4} , it is seen that s_2, s_3, \dots, s_{k+1} are dependent. Then the k points p_2, p_3, \dots, p_{k+1} congruent to them are dependent, which is not possible (*property (d)*) since these k points are contained in the $k+1$ points p_1, p_2, \dots, p_{k+1} which are independent (Theorem 6) since they belong to the pseudo- $S_{k,r}$ $(k+3)$ -tuple p_1, p_2, \dots, p_{k+3} .

Case II. The points s_2, s_3, \dots, s_{k+1} are each equidistant from the points s_{k+2}, s_{k+3} and since these k points are independent it follows from *property (k)* that the $(k-1)$ -dimensional subspace $S_{k-1,r}^*$ determined by them is the locus of points of $S_{k,r}$ equidistant from s_{k+2} and s_{k+3} . Now the points s_2, s_3, \dots, s_{k+1} are also contained in the locus of points s of $S_{k,r}$ such that $\cos(ss_{k+2}/r) + \cos(ss_{k+4}/r) = 0$. Since $s_{k+2}s_{k+4} = p_{k+2}p_{k+4}$, the points s_{k+2}, s_{k+4} are distinct and not diametral. We prove now the following assertion:

ASSERTION. *The locus of points t of $S_{k,r}$ such that*

$$\cos(ts_{k+2}/r) + \cos(ts_{k+4}/r) = 0$$

is the $(k-1)$ -dimensional subspace of $S_{k,r}$ determined by s_2, s_3, \dots, s_{k+1} .

To prove this assertion, it is shown first that if t_1, t_2, \dots, t_{k+1} are $k+1$ pairwise distinct points of $S_{k,r}$ such that $\cos(t_i s_{k+2}/r) + \cos(t_i s_{k+4}/r) = 0$, then the determinant $\Delta_{k+1}(t_1, t_2, \dots, t_{k+1})$ of these points vanishes, and hence the points are in an $S_{k-1,r}$. For suppose this determinant does not vanish, and consider $\Delta_{k+3}(t_1, t_2, \dots, t_{k+1}, s_{k+2}, s_{k+4})$, which clearly equals zero. Adding the elements of the last row (column) to the corresponding elements of the preceding row (column), and using the expansion¹⁴ employed in Theorem 2, we obtain, after some obvious reductions,

$$[1 + \cos(s_{k+2}s_{k+4}/r)] \cdot \Delta_{k+1}(t_1, t_2, \dots, t_{k+1}) = 0.$$

Since $s_{k+2}s_{k+4} \neq d$, it follows that $\Delta_{k+1}(t_1, t_2, \dots, t_{k+1})$ vanishes, contrary to our supposition. Hence each set of $k+1$ points of the locus is a dependent set, and the locus is at most $(k-1)$ -dimensional. But the locus contains the k independent points s_2, s_3, \dots, s_{k+1} of $S_{k,r}$. It follows that the locus is exactly $(k-1)$ -dimensional.

Finally, let $s \neq s_i$, ($i = 2, 3, \dots, k+1$) be any element of the $(k-1)$ -dimensional subspace determined by s_2, s_3, \dots, s_{k+1} . Now the determinant $\Delta_{k+3}(s_2, s_3, \dots, s_{k+1}, s_{k+2}, s_{k+4}, s)$ is zero, and every principal minor of order $k+2$ vanishes. It follows that every $(k+2)$ -nd order minor of the determinant is zero. Adding the $(k+2)$ -nd row (column) to the preceding row

¹⁴ See footnote 12.

(column) — a transformation of the determinant which leaves the rank unaltered — and expanding, as above, the vanishing $(k+2)$ -nd order minor $[k+2, k+2]$ of the resulting determinant, we obtain,

$$2[1 + \cos(s_{k+2}s_{k+4}/r)] \cdot \Delta_{k+1}(s_2, s_3, \dots, s_{k+1}, s) \\ - [\cos(ss_{k+2}/r) + \cos(ss_{k+4}/r)]^2 \cdot \Delta_k(s_2, \dots, s_{k+1}) = 0.$$

Since $\Delta_{k+1}(s_2, s_3, \dots, s_{k+1}, s) = 0$ and $\Delta_k(s_2, s_3, \dots, s_{k+1})$ does not vanish, we have $\cos(ss_{k+2}/r) + \cos(ss_{k+4}/r) = 0$, and s is a point of the locus. Hence the assertion is proved, and the locus in question *identified with the* $(k-1)$ -dimensional subspace $S_{k-1,r}^*$ *determined by the* k *independent points* s_2, s_3, \dots, s_{k+1} .

But this is impossible, for since $\cos(s_{k+2}s_{k+3}/r) + \cos(s_{k+3}s_{k+4}/r) = 0$, the point s_{k+3} belongs to the above locus, though it surely does not belong to $S_{k-1,r}^*$, for s_{k+3} is not equidistant from the points s_{k+2}, s_{k+4} .

This contradiction shows that the distances determined by the points $s_2, s_3, \dots, s_{k+1}, s_{k+2}, s_{k+3}, s_{k+4}$ do not satisfy the relations of Case II.

Interchanging s_{k+2} with s_{k+4} in Case III, and s_{k+3} with s_{k+4} in Case IV reduces these cases to Case II, and hence the assumption that

$$p_2, p_3, \dots, p_{k+1}, p_{k+2}, p_{k+3}, p_{k+4}$$

is not a pseudo- $S_{k,r}$ set leads to the distances determined by these points not satisfying the relations of any one of the above four cases. This contradiction proves the $k+3$ points from a pseudo- $S_{k,r}$ set. A similar procedure is used to show that the k remaining $(k+3)$ -tuples of the set P are each pseudo- $S_{k,r}$ sets, and the theorem is established.

COROLLARY. *If P is a pseudo- $S_{k,r}$ set containing more than $k+3$ points, no pair of which is diametral, then every set of $m \geq k+3$ points of P is a pseudo- $S_{k,r}$ set.*

Proof. Since P is a pseudo- $S_{k,r}$ set there is at least one pseudo- $S_{k,r}$ $(k+3)$ -tuple q_1, q_2, \dots, q_{k+3} contained in P (*property (h)*). Clearly, any subset of P that contains these $k+3$ points is a pseudo- $S_{k,r}$ set. Suppose, now, that $p_1, p_2, \dots, p_{k+3}, \dots, p_m$ is a set of $m \geq k+3$ points of P not containing any of the points q_1, q_2, \dots, q_{k+3} . Then $q_1, q_2, \dots, q_{k+3}, p_1$ is a pseudo- $S_{k,r}$ $(k+4)$ -tuple without diametral points and hence, by the preceding theorem, each set of $k+3$ of these points (in particular, the set $q_2, q_3, \dots, q_{k+3}, p_1$) is a pseudo- $S_{k,r}$ set. Then $q_2, q_3, \dots, q_{k+3}, p_1, p_2$ is a pseudo- $S_{k,r}$ $(k+4)$ -tuple without diametral points and hence, as before, $q_3, q_4, \dots, q_{k+3}, p_1, p_2$ is a pseudo- $S_{k,r}$ $(k+3)$ -tuple and $q_3, q_4, \dots, q_{k+3}, p_1, p_2, p_3$

is a pseudo- $S_{k,r}$ $(k+4)$ -tuple without diametral points. It is clear that this process may be continued until the points q_1, q_2, \dots, q_{k+3} are replaced by the points p_1, p_2, \dots, p_{k+3} forming a pseudo- $S_{k,r}$ $(k+3)$ -tuple. Then the points $p_1, p_2, \dots, p_{k+3}, \dots, p_m$ surely form a pseudo- $S_{k,r}$ set.

Finally, if $i \leq k+2$ of the points q_1, q_2, \dots, q_{k+3} occur among the m -tuple of points $p_1, p_2, \dots, p_{k+3}, \dots, p_m$, we have, with convenient labelling, $p_j = q_j$, ($j = 1, 2, \dots, i$), and the above process, starting with the pseudo- $S_{k,r}$ $(k+3)$ -tuple $q_{i+1}, q_{i+2}, \dots, q_{k+3}, p_1, p_2, \dots, p_i$, is applied as before to complete the proof of the corollary.

LEMMA. Let $P \equiv (p_1, p_2, \dots, p_{k+3}, p_{k+4})$ be a pseudo- $S_{k,r}$ $(k+4)$ -tuple without diametral points. Then $p_i p_m = p_j p_n$ or

$$\cos(p_i p_m / r) + \cos(p_j p_n / r) = 0, \\ (i, j, m = 1, 2, \dots, k+4; i \neq m \neq j).$$

Proof. If $i = j$ the lemma is trivial. Suppose, then, $i \neq j$, and consider the two $(k+3)$ -tuples

$$p_1, p_2, \dots, p_{j-1}, p_{j+1}, \dots, p_{k+4}; p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_{k+4}$$

obtained from P by omitting, in turn, the points p_j and p_i , respectively. According to Theorem 8 these two $(k+3)$ -tuples are pseudo- $S_{k,r}$ sets, and since the $(k+2)$ -tuple

$$p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_{j-1}, p_{j+1}, \dots, p_{k+4}$$

is common to both sets, an application of Theorem 7 gives at once the desired result.

THEOREM 9. Let $P \equiv (p_1, p_2, \dots, p_{k+4})$ be a pseudo- $S_{k,r}$ $(k+4)$ -tuple without diametral points. Then $p_i p_m = p_j p_n$ or

$$\cos(p_i p_m / r) + \cos(p_j p_n / r) = 0, \quad (i \neq m; j \neq n),$$

for each pair $p_i p_m, p_j p_n$ of the $\frac{1}{2}(k+3)(k+4)$ distances determined by the points of P .

Proof. If one of the indices i, m equals one of the indices j, n the theorem reduces to the preceding lemma. Suppose this is not the case. According to the lemma, $p_i p_m = p_j p_n$ or $\cos(p_i p_m / r) + \cos(p_j p_n / r) = 0$, and $p_j p_m = p_i p_n$ or $\cos(p_j p_m / r) + \cos(p_i p_n / r) = 0$. A consideration of the four possibilities thus presented leads at once to the theorem.

We may now establish a characterization theorem for pseudo- $S_{k,r}$ sets of $k+4$ points, no two points being diametral.

THEOREM 10. *If P is a pseudo- $S_{k,r}$ set of $k+4$ points, p_1, p_2, \dots, p_{k+4} , no two of which are diametral, then for every pair of distinct points p_i, p_j of P , $\cos(p_i p_j/r) = \pm 1/(k+1)$. The plus and minus signs are "determinantally distributed"; i. e., the signs occur in such a manner that the determinant*

$$\Delta_{k+4}(p_1, p_2, \dots, p_{k+4}) = |\cos(p_i p_j/r)|, \quad (i, j = 1, 2, \dots, k+4),$$

may, upon multiplication of appropriate rows and the same numbered columns by -1 , be transformed into a determinant with each element outside the principal diagonal equal to $-1/(k+1)$.

In a recent paper¹⁵ it is shown in detail that this theorem follows from Theorems 6, 7, 8, and 9 of this section, and the argument need not be repeated here.

FIRST CHARACTERIZATION THEOREM. *Let P be a pseudo- $S_{k,r}$ set of more than $k+3$ points, containing no diametral points. If p, q are any two distinct points of P , then $\cos(pq/r) = \pm 1/(k+1)$.*

Proof. If P consists of exactly $k+4$ points, the conclusion is warranted by Theorem 10. Suppose that P contains at least $k+5$ points and select any $k+4$ points of P containing the points p and q . By the Corollary to Theorem 8, these $k+4$ points form a pseudo- $S_{k,r}$ set and hence

$$\cos(pq/r) = \pm 1/(k+1).$$

LEMMA. *If p_1, p_2, \dots, p_j and q_1, q_2, \dots, q_j are two pseudo- $S_{k,r}$ sets without diametral points, and*

$$p_1, p_2, \dots, p_{j-1} \approx q_1, q_2, \dots, q_{j-1},$$

then either

$$p_i p_j = q_i q_j, \quad (i = 1, 2, \dots, j-1),$$

or

$$\cos(p_i p_j/r) + \cos(q_i q_j/r) = 0, \quad (i = 1, 2, \dots, j-1).$$

Proof. Since the two sets are pseudo- $S_{k,r}$ sets, then $j \geq k+3$. The two $(k+3)$ -tuples $p_{j-k-2}, p_{j-k-1}, \dots, p_{j-1}, p_j$ and $q_{j-k-2}, q_{j-k-1}, \dots, q_{j-1}, q_j$ are, by the Corollary to Theorem 8, pseudo- $S_{k,r}$ sets. It follows from the hypothesis of the lemma that

$$p_{j-k-2}, p_{j-k-1}, \dots, p_{j-1} \approx q_{j-k-2}, q_{j-k-1}, \dots, q_{j-1},$$

¹⁵ L. M. Blumenthal, "Metric methods in determinant theory," *American Journal of Mathematics*, vol. 61 (1939), pp. 912-922. The paper referred to uses, as noted, Theorems 6, 7, 8, 9 of the present article, but merely states these theorems without offering any proof.

and hence (Theorem 7) we have

$$(A_1). \quad p_i p_j = q_i q_j, \quad (i = j - k - 2, j - k - 1, \dots, j - 1),$$

or

$$(B_1). \quad \cos(p_i p_j / r) + \cos(q_i q_j / r) = 0, \\ (i = j - k - 2, j - k - 1, \dots, j - 1).$$

Applying the same reasoning to the two $(k + 3)$ -tuples

$$p_{j-k-3}, p_{j-k-1}, \dots, p_{j-1}, p_j \text{ and } q_{j-k-3}, q_{j-k-1}, \dots, q_{j-1}, q_j,$$

we obtain

$$(A_2). \quad p_j p_{j-k-3} = q_j q_{j-k-3}, \quad p_j p_{j-k-1} = q_j q_{j-k-1}, \dots, p_j p_{j-1} = q_j q_{j-1},$$

or

$$(B_2). \quad \cos(p_j p_{j-k-3} / r) + \cos(q_j q_{j-k-3} / r) = 0, \\ \cos(p_j p_{j-k-1} / r) + \cos(q_j q_{j-k-1} / r) = 0, \dots, \\ \cos(p_j p_{j-1} / r) + \cos(q_j q_{j-1} / r) = 0.$$

It is now easily seen that if the alternative (A_1) subsists, then the alternative (A_2) holds, while the validity of (B_1) implies that of (B_2) . Thus the alternatives (A_1) , (B_1) have been extended from $i = j - k - 2, j - k - 1, \dots, j - 1$ to $i = j - k - 3, j - k - 2, j - k - 1, \dots, j - 1$. Continuing in this manner, the index i is made to recede to 1, and the lemma is established.

SECOND CHARACTERIZATION THEOREM. *Let P be a pseudo- $S_{k,r}$ set of arbitrary power exceeding $k + 3$ and containing no diametral points. Then for every integer $i > 1$, the determinant Δ_i formed for each set of i points (pairwise distinct) of P has, upon multiplication of appropriate rows and the same numbered columns by -1 , all elements outside the principal diagonal equal to $-1/(k + 1)$.*

Proof. Let p_1, p_2, \dots, p_i be any set of $i > 1$ pairwise distinct points of P . If $i = k + 4$, the conclusion follows from Theorem 10.

Case 1. If $i < k + 4$ then the i points p_1, p_2, \dots, p_i form part of a set of $k + 4$ points which, by the Corollary to Theorem 8, is a pseudo- $S_{k,r}$ set. By Theorem 10, the determinant Δ_{k+4} of these $k + 4$ points has, upon multiplication of appropriate rows and the same numbered columns by -1 , all elements outside the principal diagonal equal to $-1/(k + 1)$. The determinant $\Delta_i(p_1, p_2, \dots, p_i)$ being a principal minor of this determinant, is then transformed by these elementary operations to the form specified in the theorem.

Case 2. If $i > k + 4$ then by using the preceding lemma in the same

manner that Theorem 7 was applied to the proof of Theorem 10, the method utilized to prove the latter theorem may be adopted *without change* to establish the present theorem in the case under consideration.

It is noted that requiring a pseudo- $S_{k,r}$ set to be free of diametral point-pairs (a condition that enters into all of the lemmas and theorems following Theorem 7) rules out pseudo- $S_{0,r}$ sets from consideration since evidently *each* pair of points of such a set has distance $d = \pi r$. It is obvious, however, that even for these sets the conclusions of the above two characterization theorems are valid.

4. Spheroidal and pseudo-spherical spaces. Characterization theorems.

A semimetric space of finite diameter d and positive space constant (parameter) ρ is called an n -dimensional spheroidal space $\mathfrak{S}_{k,\rho}^\phi$ provided Properties I-V of Section 2 are satisfied when the cosine function involved in these statements is replaced by any monotonic decreasing function $\phi(pq/\rho)$, defined for each pair of elements p, q of the space, with $\phi(0) = 1$, $\phi(d/\rho) = -1$. The $S_{n,r}$ as well as the same point set metrized with euclidean (chord) distance are examples of spheroidal spaces. It may be observed that spheroidal spaces arise as simple metric transforms of subsets of the $S_{n,r}$.

Since the derived properties (a)-(1) of Section 2 may be deduced from Properties I-V (and those properties of ϕ given above) they are valid in any spheroidal space. Thus each k -dimensional space $\mathfrak{S}_{k,\rho}^\phi$ has minimum congruence order $k + 3$ with respect to semimetric spaces. The characterization of pseudo sets of more than $k + 3$ points and free of diametral point-pairs is given by the two characterization theorems of the preceding section upon replacing the cosine function by ϕ . Thus, in particular, pseudo sets for the sphere with chord metric are characterized by these theorems.

THE UNIVERSITY OF MISSOURI,
COLUMBIA, MISSOURI.

A GEOMETRY ASSOCIATED WITH CREMONA'S EQUATIONS.*

By GERALD B. HUFF.

Introduction. In the geometry of planar Cremona transformations there are two important problems associated with the forms

$$(xx) \equiv x_1^2 + x_2^2 + \cdots + x_p^2 - x_0^2$$

and

$$(lx) \equiv x_1 + x_2 + \cdots + x_p - 3x_0.$$

If a *complete* and *regular* linear system $\Sigma_{p,d}$ of plane curves of dimension d with the generic curve having genus p is of order x_0 and has multiplicities x_1, x_2, \dots, x_ρ at a set of ρ prescribed base points, then $x \equiv \{x_0; x_1, x_2, \dots, x_\rho\}$ is called the characteristic of $\Sigma_{p,d}$ and satisfies the Cremona equations:¹

$$(xx) = 1 - d - p, \quad (lx) = -1 - d + p.$$

In 1934, Coble ² gave a method of determining every ordered solution of these equations for a given p , d , and ρ . However, a solution of the Cremona equations may not determine any system $\Sigma_{p,d}$ and there has not yet been discovered any general criterion for distinguishing between *proper*, *degenerate*, and *virtual* solutions. (A solution ³ is defined to be proper, degenerate, or virtual according as the generic curve of the system is (a) existent and irreducible, (b) existent and reducible, or (c) non-existent.) Coble gave criteria for $\rho = 9, 10$ and certain p, d but a general criterion is still lacking.

The second important problem in connection with (lx) and (xx) arises from the fact that a linear transformation,

[illegible]

which gives the effect on x of a Cremona transformation with F -points at the base points of $\Sigma_{p,d}$, must leave (xx) and (lx) absolutely invariant. Once

* Received September 13, 1939.

¹ A. B. Coble, "Algebraic geometry and theta functions," *American Mathematical Society Colloquium Publications*, vol. 10 (1929), New York City.

² A. B. Coble, "Cremona's diophantine equations," *American Journal of Mathematics*, vol. 56 (1934), pp. 459-489.

³ *Loc. cit.*, p. 461.

again, the converse is not true. There are linear transformations of this form, leaving (lx) and (xx) absolutely invariant, which do not represent any C. T.⁴

For $\rho \leq 8$, the number of these linear transformations is finite, and they all represent Cremona transformations with 8 or less F -points. For $\rho = 9$, the number is infinite but in 1932 Dr. Taylor⁵ showed that there are a finite number of types, each expressible in terms of certain parameters, and that they are all *geometric*; i. e. they all represent Cremona transformations. These results were later put in better form by Barber.⁶

For $\rho > 9$, there is still no simple way of distinguishing a geometric linear transformation. At one time it was thought that it was sufficient that the numbers n, r_i, s_j, α_{ij} be positive or zero, but examples have been devised which show that this is not true.⁴

In this paper the problem is studied by considering $(xx) = 0$ and $(lx) = 0$ as loci in a projective space S_ρ . The most interesting result is the appearance of linear transformations of infinite period and simple algebraic properties. These transformations give a simple tool for obtaining results already known and also provide the answers to questions that have been raised in the literature.

In the work the C -, P -, and D -characteristics (i. e. characteristics of (xx) , $(lx) = -1, -3; 1, -1$; and $2, 0$ respectively) play their usual important role. Also elliptic characteristics of $d = 0, p = 1$ and $d = 1, p = 2$ enter in the work for the first time. The invariant characteristic $\{3, 1^2\}$ will be designated by l and the fundamental P -characteristic $\{0; 0^{\rho-1} - 1\}$ by δ .

The group of linear transformations leaving (xx) , (lx) invariant will be denoted by $G(R)_{\rho,2}$, $G(I)_{\rho,2}$, or $G(C)_{\rho,2}$ according as the elements have rational coefficients, integer coefficients, or represent Cremona transformations. The set of ρ points at which Σ is defined will be designated by $P_{\rho,2}$.

1. Harmonic Perspectivities. The harmonic perspectivity in any point y not on $(xx) = 0$ and its polar $S_{\rho-1}: (yx) = 0$ has the equations:

$$(1) \quad x' = (yy)x - 2(yx)y.$$

⁴G. B. Huff, "A note on Cremona transformations," *Proceedings of the National Academy of Sciences*, vol. 20 (1934), pp. 428-430.

⁵M. E. Taylor, "A determination of the types of planar Cremona transformations with not more than nine F -points," *American Journal of Mathematics*, vol. 54 (1932), pp. 123-128.

⁶S. F. Barber, "Planar Cremona transformations," *American Journal of Mathematics*, vol. 56 (1934), pp. 109-121.

It will, of course, send points on $(xx) = 0$ into points on $(xx) = 0$. If it is to do the same with points on $(lx) = 0$, then either $y = l$ or $(ly) = 0$.

In the first case the equations of the involution in the point l are:

$$(2) \quad x' = (p-9)x - 2(lx)l.$$

It is readily shown that if this substitution is to leave (xx) , (lx) absolutely invariant, it must be written in the form

$$(3) \quad x' = -x + \frac{2(lx)}{p-9}l.$$

For any value of $p \neq 9$ this gives an involution which is an invariant member of $G(R)_{p,2}$. Some of these have been noticed in the literature. $p = 7, 8, 10, 11$ give elements of $G(I)_{p,2}$ which are in $G(C)_{p,2}$ for $p = 7, 8$. The involutions for $p = 3, 6, 12, 15$ have integer values for n, r_i, s_j and have been studied also.

If y is in $(lx) = 0$, the equations of the involution must be written in the form

$$(4) \quad x' = x - \frac{2(yx)}{(yy)}y$$

to insure invariance of (xx) , (lx) . For $(yy) = 2$, these are all elements of $G(I)_{p,2}$ and are the involutions in D -conditions which have been studied intensely. If y is a geometric D -condition, the involution is in $G(C)_{p,2}$.

For $(yy) = -2$, we have members of $G(I)_{p,2}$ which are never in $G(C)_{p,2}$ since $n = 1 - y_0^2$ is never a positive integer for $y_0 \neq 0$. The first integer y such that $(yy) = -2$ occurs for $p = 11$ and gives an interesting element of $G(I)_{11,2}$.

2. Pencils of Characteristics on a line. Related characteristics. We will say that two characteristics x, y are of the same sort if $(lx) = (ly)$ and $(xx) = (yy)$. This means that the associated linear systems, if any, will have the same genus p and dimension d . We may ask ourselves: under what conditions is

$$(5) \quad z = \lambda x + \mu y$$

of the same sort as both x and y ? Substitution yields:

$$(6) \quad z \equiv \lambda x + \mu y \text{ is of the same sort as both } x, y \text{ if and only if } (xy) = (xx) = (yy) \text{ and } \lambda + \mu = 1.$$

If two characteristics of the same sort satisfy the condition $(xx) = (yy) = (xy)$ we will say that they are *related*.

(7) If x and y are two related characteristics, all characteristics in the form

$$\lambda x + (1 - \lambda)y \equiv \lambda(x - y) + y$$

or

$$(1 - \mu)x + \mu y \equiv \mu(y - x) + x$$

are of the same sort as x, y .

Linear pencils of this type have been studied in detail in the literature. All D -conditions on $P_{9,2}$ are included in 120 such pencils. No related pairs of geometric P -curves or C -nets exist. Indeed, the condition that two be related is invariant under $G(C)_{p,2}$ and it is evident that $\{0; 0^{p-1}\}$ and $\{1; 0^p\}$ are not related to any geometric characteristics of the same sorts. Non-geometric pencils of related P -curves have been exhibited.⁷

If x and y of the same sort are related, the line joining x, y is tangent to $(xx) = 0$ at a point in $(lx) = 0$. Hence the result given in (7) may be put in the slightly different form:

(8) Characteristics of the form $ka + y$ are of the same sort as y for all values of k if and only if $(aa) = (la) = (ay) = 0$.

The pencils determined in (8) are the same as those given in (7). The difference is that in (7) we think of the pencil as determined by two characteristics of the same sort; and in (8) the pencil is determined by an elliptic characteristic a and a point y in the tangent S_{p-1} of $(xx) = 0$ at a .

3. Systems of characteristics which lie on a plane conic. The pencils of characteristics obtained in 2 contained characteristics lying on a line determined by two points. We can obtain systems lying on plane conics by seeking the conditions on a and b that

$$k^2a + kb + x$$

shall be of the same sort as x for all values of k . The equations

$$(l, k^2a + kb + x) \equiv (lx)$$

$$(k^2a + kb + x, k^2a + kb + x) \equiv (xx)$$

regarded as identities in k yield:

$$(\alpha) \quad (la) = (lb) = 0, \quad (aa) = (ab) = 0,$$

and

$$(\beta) \quad 2(ax) + (bb) = 0, \quad (bx) = 0.$$

If \bar{a} and \bar{b} are any two characteristics which satisfy (α) , then

⁷ See reference 2, p. 480.

$$a = \lambda \bar{a}, \quad b = \mu \bar{a} + \nu \bar{b}$$

will also satisfy (α). Substituting these in (β) gives

$$2\lambda(\bar{a}x) + \nu^2(\bar{b}\bar{b}) = 0, \quad \mu(\bar{a}x) + \nu(\bar{b}x) = 0.$$

From the second we must have

$$\mu = -\sigma(\bar{b}x), \quad \nu = \sigma(\bar{a}x).$$

Substituting in the first gives

$$2\lambda(\bar{a}x) + \sigma^2(\bar{a}x)^2(\bar{b}\bar{b}) = 0.$$

If $(\bar{a}x) = 0$, this is identically satisfied, $\nu = 0$, and we have a pencil of the type in (8). For $(\bar{a}x) \neq 0$, we must have $\lambda = -\frac{1}{2}(\bar{b}\bar{b})\sigma^2(\bar{a}x)$. (It is readily verified that if \bar{b} is an integer characteristic then $\frac{1}{2}(\bar{b}\bar{b})$ is an integer.) Thus,

$$(9) \quad \text{If } a \text{ and } b \text{ satisfy } (aa) = (la) = (lb) = (ab) = 0 \text{ then all characteristics} \\ x' = -\frac{1}{2}(\bar{b}\bar{b})(\sigma k)^2(ax)a + \sigma k[(ax)b - (bx)a] + x$$

are of the same sort as x . Moreover, every system of characteristics of the same sort and of the form

$$k^2a + kb + x$$

can be obtained in this way. If $(ax) = 0$, the system lies on the line $ka + x$.

This means that any elliptic point a and a point b on its tangent plane and $(lx) = 0$ determine for any characteristic x a system of characteristics of the same sort as x . If $(ax) \neq 0$, this system lies on a conic in the plane determined by a , b , and x . In the next paragraph we will find interesting properties of these systems.

4. The linear substitution $S_{a,b}^k$. In the preceding section it was found that any elliptic characteristic a and a characteristic b satisfying $(ab) = (lb) = 0$ determine with any characteristic x a system of characteristics of the same sort as x and lying in the plane determined by a , b , x . This led to the equation:

$$(10) \quad x' = -\frac{k^2(\bar{b}\bar{b})}{2}(ax)a + k[(ax)b - (bx)a] + x.$$

For a , b and k given, this is a linear substitution which sends any characteristic x into a characteristic x' of the same sort. It leaves (xx) , (lx) invariant. For rational a , b and k it is always an element of $G(R)_{p,2}$; for

integer a , b , and k it is in $G(I)_{p,2}$ and is sometimes an element of $G(C)_{p,2}$. We will designate such a substitution by $S^k_{a,b}$ and study its properties.

The following properties are readily verified:

$$(11) \quad \begin{aligned} S^0_{a,b} &= S^k_{a,a} = S^k_{0,b} = S^k_{a,0} = I \\ S^k_{a,b} &= S^{1ka,b} = S^{1a,kb} \\ S^{k_1}_{a,b} \cdot S^{k_2}_{a,b} &= S^{k_1+k_2}_{a,b}; \quad S^{k_1}_{a,b_1} \cdot S^{k_2}_{a,b_2} = S^{1a,k_1b_1+k_2b_2}. \end{aligned}$$

The parameter k plays the role of an exponent and gives a simple law of multiplication for substitutions $S^k_{a,b}$ defined for the same elliptic characteristic a .

The simple algebraic form of (10) leads to the following theorems:

(12) *All lines through a which are tangent to $(xx) = 0$ are left invariant by $S^k_{a,b}$ for all b, k .*

(13) *$S_{a,b} = S_{\bar{a},\bar{b}}$ if and only if $a = \lambda\bar{a}$ and $\lambda b = \mu a + \bar{b}$. That is, two substitutions are the same only if they are defined at the same point a and b, \bar{b} lie on a line through a .*

(14) *If x, y are two characteristics of the same sort, and an elliptic characteristic a exists such that*

$$(ax) = (ay) = k \neq 0,$$

then $b = (y - x)/k$ satisfies $(ab) = 0$ and defines an $S_{a,b}$ which sends x into y .

A given elliptic characteristic a and a suitable b determine a linear substitution $S_{a,b}$ and its inverse $S_{a,-b}$ which generate an infinite "cyclic" group. On the other hand, a particular elliptic characteristic a defines an aggregate of elements $S_{a,b}$ for all characteristics b satisfying

$$(ab) = (lb) = 0.$$

From the laws (11) it is evident that this aggregate constitutes an infinite abelian group. We will designate it by $\{S_a\}$.

5. The condition that $I_c I_d$ shall be a substitution $S_{a,b}$. To relate transformations $S_{a,b}$ to known elements of $G(C)_{p,2}$ we investigate the conditions under which $I_c I_d$, the product of two harmonic perspectivities in D -conditions, may be such a transformation. As an algebraic tool we use the theorem.⁸

The necessary and sufficient condition that a square matrix M can have

⁸ Given in a paper by the author read to the Texas Section of the Association, May, 1936.

its k -th power a matrix whose elements are polynomials of degree n in k is that $(M - I)^{n+1} = 0$.

Obviously, the k -th power of the matrix of any transformation $S_{a,b}$ has elements which are quadratics in k . Then if $I_c I_d$ is to be such a transformation, its matrix must satisfy $(M - I)^3 = 0$. Applying this condition, we find that we must have $(cd)^2 = 4$ which means that the line joining c and d is tangent to $(xx) = 0$ at the elliptic point $c - d$. That the condition is sufficient is shown by verifying that $S_{c-d,c} = I_c I_d$ if $(cd) = 2$.⁹

(15) If c, d are two distinct D -conditions, then $I_c I_d$ is a linear substitution $S_{a,b}$ if and only if $(cd)^2 = 4$. If signs are chosen so that c, d are related (i. e. $(cd) = 2$), then

$$I_c I_d = S_{c-d,c}$$

and $I_c I_d$ and $I_d I_c$ generate an infinite cyclic group.

In 1933 Dr. Barber,⁶ using purely algebraic methods, obtained a set of necessary and sufficient conditions that $I_c I_d$ and $I_{\bar{c}} I_{\bar{d}}$ be permutable. From the present geometrical point of view it is clear that permutability is possible if either:

- (a) the line $c\bar{d}$ is in the polar S_{p-2} of the line cd with respect to $(xx) = 0$;
- or
- (b) the lines cd and $c\bar{d}$ are tangent to $(xx) = 0$ at the same point.

Indeed, in the second case $I_c I_d$ and $I_{\bar{c}} I_{\bar{d}}$ are transformations $S_{a,b}$ defined at the same elliptic point and permutable by (11). The algebraic conditions for these two cases are equivalent to Barber's conditions and hence are necessary as well as sufficient. Thus

(16) The necessary and sufficient conditions that $I_c I_d$ and $I_{\bar{c}} I_{\bar{d}}$ be permutable is that either:

- the lines cd and $c\bar{d}$ be conjugate with respect to $(xx) = 0$,
- or the lines cd and $c\bar{d}$ be tangent to $(xx) = 0$ at the same point.

A simple algebraic form is:

(17) The necessary and sufficient condition that $I_c I_d$ and $I_{\bar{c}} I_{\bar{d}}$ be permutable is that either

⁹ $(cd) = 2$ is equivalent to $(cd)^2 = 4$ since $-c$ determines the same harmonic perspectivity and the same D -condition as c .

$$(c\bar{c}) = (c\bar{d}) = (d\bar{c}) = (d\bar{d}) = 0$$

or

$$(cd)^2 = (\bar{c}\bar{d})^2 = 4 \quad \text{and} \quad c - d = k(\bar{c} - \bar{d}),$$

where signs are chosen so that $(cd) = (\bar{c}\bar{d}) = 2$.

These conditions are simpler than those given by Barber. However, all conditions given there are necessary in a technical sense, because they are all consequences of these.

6. Results in the case $\rho = 9$. For $\rho = 9$ the S_s , $(lx) = 0$ is tangent to $(xx) = 0$ at the elliptic characteristic l . Any D -condition is on $(lx) = 0$ and by (8) defines a pencil of characteristics $kl + d$. These lines all meet $x_9 = 0$ in points which are D -conditions on $P_{9,2}$. Thus the lines of D -conditions determined by the D -conditions for which $x_9 = 0$ contain all D -conditions. There are 120 of these.

(18) All D -conditions for $P_{9,2}$ lie on the 120 tangent lines

$$kl + d$$

where d is any one of the 120 D -conditions for which $x_9 = 0$.

The one elliptic characteristic l defines a group $\{S_l\}$. Any element is determined by a characteristic b satisfying $(lb) = 0$. If one choice b defines an element, then by (13) all $\bar{b} = kl + b$ define the same element. In particular, for $k = -b_9$, there is a \bar{b} in $x_9 = 0$ which defines the element and only one of this sort. Every element of $\{S_l\}$ is determined once and only once by all characteristics b satisfying $(lb) = b_9 = 0$. Indeed, it is the infinite abelian sub-group a_9 .

(19) $\{S_l\}$ is the infinite abelian sub-group a_9 and each element is given once and only once by $S_{l,b}$, where $(lb) = b_9 = 0$.

Any two P -characteristics y, z satisfy $(ly) = (lz) = -1$, and hence by (14) $(z - y)/-1 = y - z$ is a b such that $S_{l,b}$ sends y into z .

(20) All P -characteristics on $P_{9,2}$ are geometric and any pair defines a unique element of a_9 which sends one into the other. The images of $\{0; 0^8 - 1\}$ under a_9 include all P -characteristics once and only once.

a_9 is the integer group defined at l . If we allow rational values of b then we have a subgroup of $G(R)_{9,2}$. Under this group all C -characteristics are conjugate. Indeed, two C -characteristics y, z satisfy $(ly) = (lz) = -3$ and by (14) $(y - z)/3$ is a rational b such that $S_{l,b}$ sends y into z . Ordinarily

(22) All integer P -characteristics on $P_{10,2}$ are given by

$$p = ka - l$$

where k is any integer and a is any elliptic characteristic.

Now Coble showed ² that any one of these could be reduced under $G(C)_{10,2}$ to an irreducible P -characteristic of the form $k(l + \delta) - l$, where δ is the fundamental P -characteristic $\{0; 0^9 - 1\}$ and $(l + \delta) = \{3; 1^9 0\}$ is the earliest geometric elliptic characteristic. It follows then that all elliptic characteristics are conjugate under $G(C)_{10,2}$ to $l + \delta$ or a multiple of $l + \delta$. In particular,

(23) All elliptic characteristics of positive order and $G.C.D = 1$ are geometric and reducible under $G(C)_{10,2}$ to $l + \delta = \{3; 1^9 0\}$.

Combining this with (22) yields the result

(24) All geometric P -characteristics are given by

$$p = a - l,$$

where a runs over all geometric elliptic characteristics.

That is, on each line of the cone of all P -characteristics there lies one and only one geometric characteristic. That (24) gives a definite determination of all geometric P -characteristics is clear when we recall that all elliptic characteristics are easily obtained by Coble's method.

At the particular elliptic characteristic $a = l + \delta$ there is defined a group $\{S_a\}$ which is simply isomorphic to a_9 . At any other elliptic characteristic \bar{a} there is defined a group $\{S_{\bar{a}}\}$, which is simply a transform of $\{S_a\}$ under $G(C)_{10,2}$, for a and \bar{a} are conjugate under this group.

(25) Every infinite abelian sub-group of type $\{S_a\}$ generated by pairs of involutions in related D -conditions is geometric and indeed is simply isomorphic to a_9 . Such a sub-group is defined for each elliptic characteristic and all such sub-groups are obtainable in that way.

Dr. Barber obtained by experiment one of these defined at $\{4; 2^2 1^8\}$. (25) gives a complete classification of sub-groups of this sort.

l plays a particular role in another sense. By (3), 1, the harmonic perspectivity defined by l and (lx) is a member of $G(I)_{10,2}$ and has the equations

$$T_{10}: x' = -x + 2l(lx).$$

Coble discovered this involution in another way and named it T_{10} . For T_{10} , $\{n, r_i\} = \{n; s_j\} = \{-19; -6^{10}\}$ and the P -characteristics are of the form $-2(l + \delta) + \delta$. That is, all its P -characteristics are irreducible.² This naturally raises the question: is T_{10} the only element of $G(I)_{10,2}$ with this property. Coble's simple algebraic form for irreducible P -characteristics provides an easy answer, since two P -characteristics must satisfy $(p\bar{p}) = 0$. If

$$\{3k; k^8, -1, k\} \quad \text{and} \quad \{3k'; k'^8 k' - 1\}$$

are to satisfy this, then $kk' + k + k' = 0$ or $(k + 1)(k' + 1) = 1$, which is true for integers only when $k = k' = 0, -2$. Hence,

(26) Any element of $G(I)_{10,2}$ such that all its P -characteristics are irreducible under $G(C)_{10,2}$ is either T_{10} or T_{10} multiplied by an element of π .

An important corollary is:

(27) $G(I)_{10,2}$ is generated by π , A_{123} , and T_{10} .

Another result is interesting to the writer in that it enables him to answer a question that has been in his mind for some time. It is known that $n, s_i, r_j, \alpha_{ij} \geq 0$ is not a sufficient condition that an element of $G(I)_{\rho,2}$ be in $G(C)_{\rho,2}$, the first example being devised for $\rho = 11$. However, the condition is sufficient for $\rho \leq 9$. The only case in doubt has been $\rho = 10$. From (26) we see that any element in $G(I)_{10,2}$ but not in $G(C)_{10,2}$ can be written in the form ET_{10} where E is an element of $G(C)_{10,2}$. By the algebraic form of T_{10} it can be shown that the n of such an element must be negative.

(28) For $\rho \leq 10$ the elements of $G(I)_{\rho,2}$ which have $n, r_i, s_j, \alpha_{ij} \geq 0$ are all in $G(C)_{10,2}$. For $\rho \geq 11$, this is no longer true.

8. The case $\rho = 11$. For $\rho = 11$, l defines an element of $G(I)_{11,2}$,

$$T_{11}: x' = -x + l(lx)$$

for which $\{n; s_j\} = \{-10; -3^{11}\}$. It sends any C or P -characteristic of positive order into one of negative order. Also, this is the first place a transformation I_ν is defined for $(ly) = 0$, $(yy) = -2$. The first case occurs for $\nu = \{4; 2 \cdot 1^{10}\}$ and has the equations

$$I_\nu: x' = x + \nu(\nu x).$$

Since $(l\nu) = 0$, T_{11} and T_ν are permutable and $T_{11}T_\nu$ is an involution. Indeed, it is the de Jonquieres involution defined by the geometric net $\{6; 5 \cdot 1^{10}\}$. This

is drawn to pencils of elliptic curves of genus 2, which define these virtual involutions. Could it be that these virtual involutions may have some geometrical meaning?

The study of the systems of characteristics lying on a line and on a plane conic is important in that it leads to the linear substitution $S_{a,b}$. The writer feels that the exceedingly simple laws which these satisfy should throw considerable light on the structure of the groups for $\rho \geq 11$. In particular, Theorem (14) furnishes a simple sufficient condition that two characteristics be conjugate under $G(I)_{\rho,2}$ and for $\rho = 9$ gives very simple results. The unity of the work would be increased if a simple geometrical definition of these transformations could be given.

All geometrical P -characteristics for $P_{9,2}$ occur once and only once among the P -curves of the sub-group $\{S_i\}$ for $\rho = 9$. From examples studied for larger ρ it seems possible that the aggregate of all geometric subgroups $\{S_a\}$, defined at all geometric elliptic characteristics a might have this property. An affirmative answer to this conjecture would make the study of $G(C)_{\rho,2}$ dependent only on the nature of the elliptic characteristics defined for that value of ρ . Thus elliptic characteristics may play as important a role in the general theory as C , P , and D -characteristics.

SOUTHERN METHODIST UNIVERSITY,
DALLAS, TEXAS.

POLYNOMIALS WHOSE REAL PART IS BOUNDED ON A GIVEN CURVE IN THE COMPLEX PLANE.*

By A. C. SCHAEFFER and G. SZEGÖ.

Introduction. 1. In what follows we denote a rational polynomial of the complex variable z by π_n if the degree of this polynomial is n .

As a simple consequence of the theorem of S. Bernstein on trigonometric polynomials, the following holds: ¹

A. Let $f(z)$ be a π_n and $|f(z)| \leq 1$ in $|z| \leq 1$. Then $|f'(z)| \leq n$ in $|z| \leq 1$, with the equality only if $f(z) = \epsilon z^n$, $|\epsilon| = 1$.

This theorem has been generalized by Szegö in two different directions. First, the unit circle may be replaced by a Jordan curve subject to certain restrictions: ²

B. Let Γ be an open or a closed Jordan curve consisting of a finite number of analytic arcs which join so that the exterior angle ³ is greater than zero. If $f(z)$ is a π_n satisfying $|f(z)| \leq 1$ on Γ , then at any point z_0 of Γ

$$|f'(z_0)| \leq An^a.$$

Here A is a constant which depends only on Γ and z_0 , and $\alpha\pi$ is the exterior angle of Γ at z_0 . The order of this bound as n becomes infinite can not be improved.

On the other hand, the condition $|f(z)| \leq 1$ in Theorem A can be replaced by $|\Re f(z)| \leq 1$, so the following is true: ⁴

C. Let $f(z)$ be a π_n and $|\Re f(z)| \leq 1$ in $|z| \leq 1$. Then $|f'(z)| \leq n$ in $|z| \leq 1$, with the equality only if $f(z) = \epsilon z^n$, $|\epsilon| = 1$.

* Received April 22, 1940.

¹ M. Riesz, "Eine trigonometrische Interpolationsformel und einige Ungleichungen für Polynome," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 23 (1914), pp. 354-368. See also, O. Szász, "Korlátos hatványsorokról," *Mathematikai és Természettudományi Értesítő*, vol. 43 (1926), pp. 504-520.

² G. Szegö, "Über einen Satz von A. Markoff," *Mathematische Zeitschrift*, vol. 23 (1925), pp. 45-61.

³ In case Γ is a closed Jordan curve, the exterior angle at any point of Γ is defined as usual. If Γ is an open curve, the exterior angle is defined as in *loc. cit.*,² pp. 48-49.

⁴ G. Szegö, "Über einen Satz des Herrn Serge Bernstein," *Königsberger Gelehrte Gesellschaft, Naturwissenschaftliche Klasse*, 1928, pp. 59-70. Also, S. Bernstein, "Sur un théorème de M. Szegö," *Prace Matematyczno-Fizyczne*, vol. 44 (1937), pp. 9-14.

2. The main result of the present note is:

THEOREM 1. *Let Γ be a closed Jordan curve consisting of a finite number of analytic arcs which join in such a way that the exterior angle is always greater than zero and less than 2π . Let $f(z)$ be a π_n satisfying*

$$(1) \quad |\Re f(z)| \leq 1, \quad z \in \Gamma.$$

Then at an arbitrary point z_0 of Γ

$$(2) \quad |f'(z_0)| \leq An^\alpha;$$

here A is a constant which depends only on Γ and z_0 , and $\alpha\pi$ is the exterior angle of Γ at z_0 . The order of this bound as n becomes infinite can not be improved.

This is a generalization of Theorem C, at least so far as the order of the bound of $|f'(z_0)|$ is concerned; for if Γ is a circle, the exterior angle at every point is π so that $\alpha = 1$. Incidentally, in this special case our general method used in § 2, furnishes the inequality

$$|f'(z)| < 6en, \quad |z| \leq 1.$$

For closed Jordan curves, Theorem 1 is an obvious extension of Theorem B which was obtained under the more restrictive hypothesis $|f(z)| \leq 1$ on Γ . Theorem B, however, holds even if Γ is an open arc, while our Theorem 1 does not. Indeed let Γ be the real segment $-1 \leq x \leq +1$ and consider the polynomial $f(z) = iKz$, $z = x + iy$, K real. In this case $\Re f(z) = 0$ on Γ , but $|f'(1)|$ can be arbitrarily large. More generally, we can take for Γ a Jordan arc along which the real part of a certain given polynomial is constant.

Our proof of Theorem 1 makes use of the theory of conformal mapping and in particular of the theorems of Osgood-Taylor⁵ concerning the behavior of the map-function near the boundary. (See however, the last remark in § 2.)

3. Under the conditions of Theorem 1 we may ask for proper bounds for the "oscillation" of $f(z)$ in Γ , that is for the maximum of $|f(z_1) - f(z_2)|$ if z_1 and z_2 describe, independently of each other, the closed interior of Γ .

THEOREM 2. *Let Γ have the same meaning as in Theorem 1 and let $f(z)$ satisfy the same conditions as there. Then for two arbitrary points z_1 and z_2 in the closed interior of Γ ,*

$$(3) \quad |f(z_1) - f(z_2)| < A \log n, \quad n > 1;$$

⁵ W. F. Osgood and E. H. Taylor, "Conformal transformations on the boundaries of their regions of definition," *Transactions of the American Mathematical Society*, vol. 14 (1913), pp. 227-228.

here A depends only on Γ . The order of this bound as n becomes infinite can not be improved.

Let $f(z)$ be real at a certain (not necessarily fixed) point in Γ ; then from (3)

$$(3') \quad |\Im f(z)| < A \log n, \quad z \in \Gamma,$$

follows.

Theorem 2 is well-known for the case in which Γ is the unit circle. The much discussed example

$$f(z) = (i/2)(z/1 + z^2/2 + \cdots + z^n/n)$$

shows that $\log n$ is the true rate of growth of the bound in (3) or (3') [$f(0) = 0$] in case Γ is the unit circle.

Theorem 2 has a more elementary character than Theorem 1; therefore we found it convenient to bring its proof first. Having proved both inequalities, we discuss the precision of our estimates as n becomes infinite. Obviously Theorem 2 combined with Theorem B furnishes the less informative bound $An^a \log n$ instead of the bound in (2).

In the proofs of both theorems we use the following

LEMMA 1. *Let Γ satisfy the conditions of Theorem B and let $f(z)$ be a π_n satisfying $|f(z)| \leq 1$, $z \in \Gamma$. We denote by $\psi(z)$ a function which maps the exterior of Γ onto the exterior of the unit circle in such a way that the points at infinity correspond. Then at a point z' outside Γ*

$$(4) \quad |f(z')| \leq |\psi(z')|^n.$$

Here $|\psi(z')| = R > 1$.

This is a well-known consequence of the maximum principle.

1. Proof of Theorem 2. 1. Let Γ satisfy the conditions of Theorem 2, and let $\beta > 0$ be the smallest interior angle at which two arcs of Γ join. If z_0 is any point on Γ , we draw through z_0 two line segments L_1 and L_2 with the following properties:

- (a) z_0 is one end-point of L_1 and of L_2 ;
- (b) the other end-points of L_1 and L_2 are also on Γ , whereas all other points of L_1 and L_2 are in the open interior of Γ ;
- (c) at z_0 , L_1 and L_2 intersect Γ (or one of the arcs of Γ if z_0 is a vertex) with an angle of $\beta/4$.

The distance of any point Z on L_1 or L_2 from Γ (that is, from any point z on Γ) is at least $\sin(\beta/8)$ times the distance from Z to z_0 provided Z is

sufficiently near to z_0 . We determine the *largest* segments L'_1 and L'_2 on L_1 and L_2 respectively, having z_0 as one end-point, for whose points Z the condition

$$(5) \quad |Z - z| \geq |Z - z_0| \sin(\beta/8), \quad z \in \Gamma,$$

is satisfied. In what follows, let $L = L(z_0)$ denote the larger of the segments L'_1 and L'_2 (or either of them if they are equal). The length $l(z_0)$ of this $L(z_0)$ has a *positive minimum*, say l_0 , as z_0 runs over Γ .

2. The following statements are essentially known:

LEMMA 2. *If the real part of an analytic function $F(z)$ is bounded by 1 in the open interior of a circle of radius $r > 0$, then at the center z' of this circle*

$$(6) \quad |F'(z')| \leq 2/r.$$

LEMMA 3. *If the real part of an analytic function $F(z)$ is bounded by 1 in the unit circle $|z| < 1$, then*

$$(7) \quad |F(z) - F(0)| \leq 2 \log \frac{1}{1 - |z|}.$$

LEMMA 4. *Let S be a segment of length s . If $f(z)$ is a π_n satisfying $|f(z)| \leq 1$ on S , then $|f(z)| \leq K$ provided z lies within a distance n^{-2} of either end-point of S . Here K is a constant which depends on s but is independent of n .*

Inequality (6) may be obtained by differentiating Poisson's integral which for a circle of radius ρ , $\rho < r$, with center at the origin is

$$F(z) = i \cdot \Im\{F(0)\} + \frac{1}{2\pi} \int_{-\pi}^{+\pi} \frac{\rho e^{i\phi} + z}{\rho e^{i\phi} - z} \cdot \Re\{F(\rho e^{i\phi})\} d\phi, \quad |z| < \rho.$$

Inequality (7) may be obtained from (6) by integrating along a radius from 0 to z :

$$|F(z) - F(0)| = \left| \int_0^z F'(\xi) d\xi \right| \leq \int_0^{|z|} \frac{2}{1-t} dt$$

which is (7). Lemma 4 follows from Lemma 1 of the Introduction where $\psi(z)$ is a function which maps the exterior of S onto the exterior of the unit circle with the points at infinity corresponding. In case S is the segment $(-1, +1)$ of the real axis we need only note that $|\psi(z)|^n = |z + (z^2 - 1)^{1/2}|^n$ is bounded if $|z - 1| \leq n^{-2}$ or $|z + 1| \leq n^{-2}$.

3. Now we proceed to the proof of Theorem 2. Let ξ_0 be a fixed interior point of Γ and $F(z)$ an analytic function (not necessarily a polynomial) satisfying the condition $|\Re F(z)| \leq 1$, $z \in \Gamma$. Then, according to Lemma 3,

$$(8) \quad |F(z) - F(\zeta_0)| \leq 2 \log \frac{1}{1 - |\phi(z)|}, \quad z \in \Gamma$$

where $w = \phi(z)$ is a function which maps the interior of Γ onto the circle $|w| < 1$ such that $\phi(\zeta_0) = 0$. If δ is a small positive number, let A_δ be the set of points inside Γ which lie at a distance δ or greater from Γ . Let $\delta = l_0 \sin(\beta/8)$; then the segments L drawn through each point z_0 of Γ according to the former construction, extend into A_δ . Furthermore, from (8),

$$(9) \quad |F(z) - F(\zeta_0)| \leq B, \quad z \in A_\delta;$$

here B is a constant which depends only on Γ and ζ_0 . Now let z_0 be a point on Γ and let z_1 be the end-point of L different from z_0 ; then $z_1 \in A_\delta$. If ζ is any point of L , (5) and (6) imply that

$$|F'(\zeta)| \leq 2\{|\zeta - z_0| \sin(\beta/8)\}^{-1}.$$

This, together with (9), shows that if z is any point on L ,

$$(10) \quad |F(z) - F(\zeta_0)| \leq B + \left| \int_{z_1}^z F'(\zeta) d\zeta \right| \leq C \log \frac{C}{|z - z_0|}$$

where C is again a positive constant which depends only on Γ and ζ_0 .

So far the polynomial character of our function has not been used. Now let $F(z) = f(z)$ be the π_n of Theorem 2. In the portion of L which lies at a distance greater than n^{-2} from z_0 , (10) shows that

$$|f(z) - f(\zeta_0)| \leq C \log(Cn^2).$$

Since the length of L is greater than a fixed positive number l_0 , Lemma 4 implies that

$$|f(z) - f(\zeta_0)| \leq KC \log(Cn^2)$$

where K depends only on Γ . This completes the proof of Theorem 2 because

$$|f(z_1) - f(z_2)| \leq |f(z_1) - f(\zeta_0)| + |f(z_2) - f(\zeta_0)|.$$

2. Proof of Theorem 1. 1. Let z_0 be a point of Γ at which two arcs γ_1 and γ_2 intersect with an exterior angle $\alpha\pi$, $0 < \alpha < 2$. It is no loss of generality to suppose that $z_0 = 0$ and that the tangents to γ_1 and γ_2 at $z_0 = 0$ intersect the real axis at angles of $\alpha\pi/2$ and $-\alpha\pi/2$, respectively; also we may assume that a neighborhood of the negative real axis near the origin lies inside Γ . With ρ a small positive number draw two circles of radius ρ , the first with center at $\rho \exp\{i(\alpha+1)\pi/2\}$ and the second with center at $\rho \exp\{-i(\alpha+1)\pi/2\}$. These two circles will intersect at the origin, where they are tangent to γ_1 and γ_2 , respectively, and at the point

$$z = 2\rho \cos \{(\alpha + 1)\pi/2\}$$

on the negative real axis. The arc of the first circle which lies above and on the real axis, and the arc of the second circle which lies beneath and on the real axis, together form the boundary of a region R which is closed and simply connected and whose boundary touches Γ at $z_0 = 0$. All other points of R will lie inside Γ if ρ is small enough, and the exterior angle of R at z_0 is $\alpha\pi$.

If $\alpha = 1$ the two arcs which form the boundary of R are arcs of the same circle, and R is the circle $|z + \rho| \leq \rho$.

If δ is a small positive number, let R_δ be the region obtained by translating R a distance δ to the left; that means $z \in R_\delta$ if and only if $(z + \delta) \in R$. Now we show the following. If ρ is small but fixed, then for all sufficiently small δ the region R_δ will lie entirely inside Γ and at a distance at least $\lambda\delta$ from Γ where λ is a positive constant independent of δ . Indeed let us repeat the previous construction of R replacing ρ by 3ρ ; the resulting region S is bounded by two arcs of circles of radius 3ρ with centers at

$$3\rho \exp \{ \pm i(\alpha + 1)\pi/2 \};$$

choose ρ so small that S lies entirely in the closed interior of Γ . Now fix ρ ; for

$$0 < \delta < \rho |\cos \{(\alpha + 1)\pi/2\}|$$

one shows by direct calculation that R_δ lies inside S and at a distance greater than

$$\frac{1}{2}\delta |\cos \{(\alpha + 1)\pi/2\}|$$

from the boundary of S , and so inside Γ and at least this distance from Γ .

2. Let $f(z)$ be a polynomial satisfying the conditions of Theorem 1. We conclude from (6) that if z is any point of R_δ

$$(11) \quad |f'(z)| \leq 2/(\lambda\delta).$$

Let $\psi(z)$ be a function which maps the exterior of R onto the exterior of the unit circle with the points at infinity mutually corresponding. Then $\psi(z + \delta)$ maps the exterior of R_δ onto the exterior of the unit circle, and we obtain from (11) by application of Lemma 1 [cf. (4)]

$$|f'(0)| \leq \{2/(\lambda\delta)\} |\psi(\delta)|^n.$$

But from a theorem of Osgood-Taylor mentioned in the Introduction [see ⁵] we conclude that near the boundary point $z = 0$ of R the map-function $\psi(z)$ must be of the form

$$\psi(z) = \psi(0) + z^{1/\alpha} p(z)$$

where $|\psi(0)| = 1$ and $p(z)$ approaches a finite limit not zero as z approaches zero. Then if $|p(z)| \leq A$ for small $|z|$, we obtain

$$|f'(0)| \leq \{2/\lambda\delta\}(1 + A\delta^{1/a})^n.$$

Placing $\delta = n^{-a}$ (permissible for large n) this gives

$$|f'(0)| < 2 \cdot \lambda^{-1} \cdot e^A \cdot n^a$$

which proves the theorem.

We notice that the map-function $\psi(z)$ of the region R may be calculated in terms of elementary functions. This makes it possible to avoid the use of the Osgood-Taylor theorem.

3. Discussion of the precise order. 1. The bound An^a in Theorem 1 is of the precise order as $n \rightarrow \infty$; this follows from the corresponding fact in Theorem B.

We show that the bound $A \log n$ in Theorem 2 is also the precise one. More exactly, let Γ_0 be a closed region in the open interior of Γ , z_1 a fixed point on Γ and z_2 arbitrary in Γ_0 ; we construct a sequence $\{g_n(z)\}$ such that $g_n(z)$ is a π_n , $n \geq 1$, and

$$\begin{aligned} |\Re g_n(z)| &\leq 1, & z \in \Gamma, \\ |g_n(z_1) - g_n(z_2)| &> A' \log n; \end{aligned}$$

here $A' > 0$ is independent of n .

2. By use of the polynomials $\sum_1^n z^{\nu}/\nu$ this construction is rather easy in case a circle through z_1 exists containing Γ . The following method holds generally. The principal tool is Faber's polynomials $f_n(z)$, $n \geq 1$, associated with Γ . They are defined as follows. Let

$$(12) \quad \begin{aligned} w = \psi(z) &= cz + c_0 + c_1 z^{-1} + \cdots, \\ z = \psi^{-1}(w) &= c^{-1}w + \cdots, & c > 0, \end{aligned}$$

be the conformal mapping of the exterior of Γ onto the exterior of the unit circle $|w| > 1$, uniquely determined by the condition $c > 0$. Then $f_n(z)$ is defined as the "principal part" of $\{\psi(z)\}^n$, that is⁶

$$(13) \quad f_n(z) = \frac{1}{2\pi i} \int_C \frac{\{\psi(Z)\}^n}{Z - z} dZ = \frac{1}{2\pi i} \int \frac{W^n \{\psi^{-1}(W)\}'}{\psi^{-1}(W) - z} dW.$$

Here the integration is extended over a curve C enclosing Γ (and over the corresponding curve in the w -plane), and z is in the interior of C . For the construction mentioned we need the following expansion (slightly different from the expansion in ², p. 54, (17)):

⁶ See ², p. 53.

$$(14) \quad \log \frac{\psi^{-1}(W) - z}{W} = \log \frac{1}{c} - \sum_{m=1}^{\infty} \frac{f_m(z)}{m W^m}.$$

Here the determination of the logarithms is obvious; and z is arbitrary. But $|W| > 1$ if z is in the interior of Γ , and $|W| > |\psi(z)|$ if z is in the exterior of Γ .

Expansion (14) is clear for $W = \infty$. The differentiated expansion

$$(15) \quad \frac{\{\psi^{-1}(W)\}'}{\psi^{-1}(W) - z} = W^{-1} + \sum_{m=1}^{\infty} f_m(z) W^{-m-1}$$

follows from (13).

3. Let z be arbitrary in the closed interior of Γ , and let $|W| > 1$. Then the imaginary part of (14) is uniformly bounded (see ², p. 54) so we have for the Cesàro means of first order

$$(16) \quad \left| \sum_{m=1}^n \left(1 - \frac{m}{n}\right) \frac{f_m(z)}{m W^m} \right| \leq Q, \quad z \in \Gamma, |W| \geq 1,$$

where Q depends only on Γ . Also, the function (14) is bounded for $|W| \geq 1$ and for a fixed z in the open interior of Γ (uniformly if z is restricted to a closed region Γ_0 entirely in the open interior of Γ); that is

$$(17) \quad \left| \sum_{m=1}^n \left(1 - \frac{m}{n}\right) \frac{f_m(z)}{m W^m} \right| \leq Q', \quad z \in \Gamma_0, |W| \geq 1,$$

where Q' depends only on Γ and Γ_0 .

4. Let z_1 be an arbitrary point on Γ with the exterior angle $\alpha\pi$, $0 < \alpha < 2$, and let $w_1 = \psi(z_1)$, $|w_1| = 1$. Assuming for a moment that Γ is a closed polygon, we find by use of the Schwarz-Christoffel formula

$$(18) \quad \psi^{-1}(W) - \psi^{-1}(w_1) = (W - w_1)^{\alpha} F(W - w_1)$$

where $F(t)$ is analytic around $t = 0$, and $F(0) \neq 0$. This furnishes, if $|W - w_1|$ is sufficiently small,

$$(19) \quad \frac{\{\psi^{-1}(W)\}'}{\psi^{-1}(W) - \psi^{-1}(w_1)} = \frac{\alpha}{W - w_1} + \frac{F'}{F}(W - w_1).$$

Now,

$$(20) \quad f_n(z_1) = \frac{1}{2\pi i} \int W^n \left\{ \frac{\{\psi^{-1}(W)\}'}{\psi^{-1}(W) - \psi^{-1}(w_1)} - \frac{\alpha}{W - w_1} \right\} dW \\ + \frac{\alpha}{2\pi i} \int \frac{W^n}{W - w_1} dW.$$

For the line of integration we choose two arcs c_1 and c_2 ; c_1 connects two points w' and w'' of the unit circle (on opposite sides of w_1) and runs entirely

in the exterior $|W| > 1$ of the unit circle; the other arc c_2 is the "large" arc $w'w''$ of the unit circle $|W| = 1$. We choose w', w'', c_1 so that the function $F(W - w_1)$ is regular and $\neq 0$ in the domain bounded by the "small" arc $w'w''$ of the unit circle and by c_1 .

In the first integral of (20) we can replace $c_1 + c_2$ by the unit circle $|W| = 1$; the resulting integral approaches 0 as $n \rightarrow \infty$, according to Riemann's lemma. The second integral is αw_1^n , so

$$(21) \quad f_n(z_1) = \alpha w_1^n + o(1), \quad n \rightarrow \infty.$$

5. We define the required polynomials by

$$g_n(z) = \frac{i}{Q} \sum_{m=1}^n \left(1 - \frac{m}{n}\right) \frac{f_m(z)}{m w_1^m}.$$

According to (16) and (17)

$$(22) \quad |\Re\{g_n(z)\}| \leq 1, \quad z \in \Gamma; \quad |g_n(z)| \leq \frac{Q'}{Q}, \quad z \in \Gamma_0.$$

But according to (21),

$$(23) \quad g_n(z_1) = i \frac{\alpha}{Q} \sum_{m=1}^n \left(1 - \frac{m}{n}\right) \frac{1}{m} + o(\log n) \\ = i \frac{\alpha}{Q} \log n + o(\log n), \quad n \rightarrow \infty.$$

This shows that the bound $A \log n$ in Theorem 2 is of the right order as $n \rightarrow \infty$.

6. Finally we remove the condition that Γ is a polygon. If z_1 is given, we construct a polygon Γ' with the following properties:

- (a) Γ' contains Γ ;
- (b) z_1 is on Γ' ;
- (c) the exterior angle of Γ' at z_1 is $\alpha'\pi$, $0 < \alpha' < 2$.

Obviously there is no difficulty in constructing such a polygon Γ' so long as $\alpha' < \alpha$.

Repeating the previous consideration for Γ' , we obtain a sequence of π_n satisfying conditions (22); instead of (23) we have

$$g_n(z_1) = i \frac{\alpha'}{Q} \log n + o(\log n), \quad n \rightarrow \infty,$$

which suffices for our purpose.

**NEUER BEWEIS EINES SATZES VON G. H. HARDY UND
S. RAMANUJAN ÜBER DAS ASYMPTOTISCHE
VERHALTEN DER ZERFÄLLUNGS-
KOEFFIZIENTEN.***

VON VOJISLAV G. AVAKUMOVIĆ.

Wird mit $p(n)$ die Anzahl der verschiedenen Zerlegungen von n in gleiche oder ungleiche positive ganzzahlige Summanden bezeichnet, so ergibt bekanntlich die Hardy-Ramanujansche asymptotische Entwicklung von $p(n)$ in erster Annäherung die Formel

$$1) \quad p(n) \sim \frac{1}{4\sqrt{3}n} \exp \left[2\sqrt{\frac{\pi^2}{6}n} \right], \quad n \rightarrow \infty.^1$$

Einen Beweis dieser Formel habe ich auf Grund allgemeiner Tauberscher Sätze funktionentheoretischer Art im Sections-Vortrag "Über das Verhalten Laplacescher Integrale an der Konvergenzgrenze u. s. w." /2. Congr. Interbalkan. des Math. Bucarest, 12-IX-1937. Bull. Math. Soc. Roum. Sci. 40, Nr. 1/2 1938, S. 101-106/ gegeben.²

Im folgenden möchte ich mit der im Prinzip gleichen Methode die Formel I) auf möglichst kurzem Wege beweisen.

1) Für $R(s) > 0$ ist

$$g(s) = \prod_{n=1}^{\infty} \frac{1}{1 - e^{-sn}} = 1 + \sum_{n=1}^{\infty} p(n)e^{-sn},$$

also,

$$(1) \quad A(u) = p(n) \quad \text{für} \quad n \leq u < n+1, \quad (n = 0, 1, 2, \dots)$$

gesetzt,

$$(2) \quad \int_0^{\infty} e^{-su} A(u) du = \left(\frac{1 - e^{-s}}{s} \right) g(s).$$

Sei

$$\phi(u) = \begin{cases} 0 & \text{für } 0 \leq u < 1/24 \\ 1 & \text{für } 1/24 \leq u < 25/24 \\ 0 & \text{für } 25/24 \leq u \end{cases}$$

* Received April 29, 1940.

¹ G. H. Hardy and S. Ramanujan, "Asymptotic formulae in combinatory analysis," *Proceedings of the London Mathematical Society* (2), vol. 17 (1918), pp. 75-115.

² Den Beweis eines Specialfalles dieser Sätze habe ich in der Note: "Théorèmes relatifs aux intégrales de Laplace sur leur frontière de convergence," *C. R. de l'Acad. des Sci. Paris*, vol. 204 (1937), pp. 224-226 skizziert.

und

$$B(u) = 1/\sqrt{2\pi} \int_0^u \phi(t) \sum_{\nu=1}^{\infty} (\pi^2/6)^\nu \frac{(u-t)^{\nu-3/2}}{\Gamma(\nu+1)\Gamma(\nu-1/2)} dt.$$

Dann ist

$$\begin{aligned} \int_0^\infty e^{-su} B(u) du \\ &= 1/\sqrt{2\pi} \int_0^\infty e^{-su} \phi(u) du \int_0^\infty e^{-su} \sum_{\nu=1}^{\infty} (\pi^2/6)^\nu \frac{u^{\nu-3/2}}{\Gamma(\nu+1)\Gamma(\nu-1/2)} du, \\ &= \left(\frac{1-e^{-s}}{s} \right) e^{-s/24} \sqrt{s/2\pi} (\exp[\pi^2/6s] - 1), \end{aligned}$$

was zusammen mit (2)

$$\begin{aligned} \int_0^\infty e^{-su} \{A(u) - B(u)\} du \\ &= \left(\frac{1-e^{-s}}{s} \right) g(s) - \left(\frac{1-e^{-s}}{s} \right) e^{-s/24} \sqrt{s/2\pi} (\exp[\pi^2/6s] - 1) \\ &= J(s) \end{aligned}$$

ergibt. Auf Grund der für die elliptische Modulfunktion $g(s)$ gültigen Funktionalgleichung

$$g(s) = \sqrt{s/2\pi} \exp[-s/24 + \pi^2/6s] g(4\pi^2/s)$$

sieht, man, dass

$$J(\epsilon + t^2 + 2ait), \quad i = \sqrt{-1}$$

bei festem a für jedes $\epsilon > 0$ eine im Intervall $(-\infty, +\infty)$ gleichmässig in ϵ beschränkte und absolut integrable Funktion darstellt. Also ist

$$\begin{aligned} a \int_{-\infty}^{+\infty} e^{2axit} J(\epsilon + t^2 + 2ait) dt \\ &= a \int_0^{+\infty} e^{-\epsilon u} \{A(u) - B(u)\} du \int_{-\infty}^{+\infty} \exp[-t^2 u + 2ai(x-u)t] dt \\ &= a\sqrt{\pi} \int_0^\infty \{A(u) - B(u)\} \exp\left[-\epsilon u - a^2 \frac{(x-u)^2}{u}\right] du / \sqrt{u}, \end{aligned}$$

woraus

$$\begin{aligned} (3) \quad a \int_0^\infty \{A(u) - B(u)\} \exp\left[-a^2 \frac{(x-u)^2}{u}\right] du / \sqrt{u} \\ &= a/\sqrt{\pi} \int_{-\infty}^{+\infty} e^{2axit} J(t^2 + 2ait) dt \end{aligned}$$

folgt, da im Integral rechts wegen $\limsup_{u \rightarrow \infty} |A(u) - B(u)| \exp[-\delta u] < \text{const.}$

(für jedes $\delta > 0$) der Grenzübergang $\epsilon \rightarrow 0$ erlaubt ist. Wegen

$$B(u) = 1/\sqrt{2\pi} \sum_{\nu=1}^{\infty} (\pi^2/6)^{\nu} \frac{(u-1/24)^{\nu-1/2} - (u-25/24)^{\nu-1/2}}{\Gamma(\nu+1)\Gamma(\nu-1/2)(\nu-1/2)}, \quad (u \geq 25/24)$$

$$\sim \frac{1}{4\sqrt{3}u} \exp \left[2\sqrt{\frac{\pi^2}{6}u} \right], \quad u \rightarrow \infty$$

ist

$$a \int_0^{\infty} B(u) \exp \left[-a^2 \frac{(x-u)^2}{u} \right] du / \sqrt{u}$$

$$\sim \frac{\sqrt{\pi}}{4\sqrt{3}x} \exp \left[\pi^2/24a^2 + 2\sqrt{\frac{\pi^2}{6}x} \right], \quad x \rightarrow \infty,$$

so dass aus (3) schliesslich

$$(4) \quad a \int_0^{\infty} A(u) \exp \left[-a^2 \frac{(x-u)^2}{u} \right] du / \sqrt{u}$$

$$\sim \frac{\sqrt{\pi}}{4\sqrt{3}x} \exp \left[\pi^2/24a^2 + 2\sqrt{\frac{\pi^2}{6}x} \right], \quad x \rightarrow \infty$$

folgt.³

2) Mit $y = x - \sqrt{x/a}$ ist

$$1 + o(1) \geq \exp[-\pi^2/24a^2] \left(x/y \exp \left[2\sqrt{\frac{\pi^2}{6}y} - 2\sqrt{\frac{\pi^2}{6}x} \right] \right)$$

$$\times 4y\sqrt{3} \exp \left[-2\sqrt{\frac{\pi^2}{6}y} \right] A(y) \frac{a}{\sqrt{\pi}} \int_y^{\infty} \exp \left[-a^2 \frac{(x-u)^2}{u} \right] \frac{du}{\sqrt{u}},$$

also

$$(5) \quad \limsup_{y \rightarrow \infty} A(y) 4\sqrt{3}y \exp \left[-2\sqrt{\frac{\pi^2}{6}y} \right]$$

$$\leq \frac{\sqrt{\pi} \exp[\pi^2/24a^2 - \sqrt{\pi^2 6a}]}{\int_{-\sqrt{a}}^{\infty} e^{-u^2} du} = W_1(a) \rightarrow 1, \quad a \rightarrow \infty.$$

3) Sei $\Omega = \Omega(t)$ die kleinste nicht-negative, nach (5) stets vorhandene Funktion, für die

$$(1 + \Omega) \frac{1}{4\sqrt{3}t} \exp \left[2\sqrt{\frac{\pi^2}{6}t} \right] - A(t) \geq 0$$

ist. Wird zur Abkürzung

$$\Omega_x = \text{Min } \Omega(t)$$

$$x - \sqrt{x/a} \leq t \leq x$$

gesetzt, so folgt wegen (4) und $\Omega \rightarrow 0, t \rightarrow \infty$

³ Bei $x \rightarrow \infty$ strebt das in (3) rechts stehende Integral als Fourierconstante einer absolut integrierbaren Funktion gegen 0.

$$\begin{aligned}
 o(1) &= x \exp \left[-2 \sqrt{\frac{\pi^2}{6}} x \right] a \int_0^\infty \left\{ (1 + \Omega) \frac{1}{4\sqrt{3} u} \right. \\
 &\quad \times \exp \left[2 \sqrt{\frac{\pi^2}{6}} u \right] - A(u) \left. \right\} \exp \left[-a^2 \frac{(x-u)^2}{u} \right] du / \sqrt{u} \\
 &\geq x \exp \left[-2 \sqrt{\frac{\pi^2}{6}} x \right] \frac{(1 + \Omega_x)}{4\sqrt{3}} a \int_{x-\sqrt{x/a}}^x \exp \left[2 \sqrt{\frac{\pi^2}{6}} u - a^2 \frac{(x-u)^2}{u} \right] du / u^{3/2} \\
 &\quad - x \exp \left[-2 \sqrt{\frac{\pi^2}{6}} x \right] A(x) a \int_{x-\sqrt{x/a}}^x \exp \left[-a^2 \frac{(x-u)^2}{u} \right] du / \sqrt{u},
 \end{aligned}$$

also

$$\begin{aligned}
 (6) \quad \liminf_{a \rightarrow \infty} A(x) 4\sqrt{3} x \exp \left[-2 \sqrt{\frac{\pi^2}{6}} x \right] \\
 \geq \frac{\int_{-\sqrt{a}}^0 \exp \left[\sqrt{\frac{\pi^2}{6}} u/a - u^2 \right] du}{\int_{-\sqrt{a}}^0 e^{-u^2} du} = W_2(a) \rightarrow 1, \quad a \rightarrow \infty.
 \end{aligned}$$

4) Aus (1), (5) und (6) folgt I).

UNIVERSITY, MATHEMATICAL SEMINAR,
BEOGRAD, YUGOSLAVIJA.

AN ALGEBRAIC PROBLEM INVOLVING THE INVOLUTORY INTEGRALS OF LINEAR DYNAMICAL SYSTEMS.*

By JOHN WILLIAMSON.

Introduction. In what follows $f = f(x)$, $g = g(x)$ etc. are scalar functions with continuous second derivatives and are not constant in the x -domain under consideration. The point $x = (x_1, x_2, \dots, x_{2n})$ is a point of the $2n$ -dimensional phase space

$$x_i = p_i, x_{n+i} = q_i, \quad (i = 1, 2, \dots, n),$$

where, with the usual notation of dynamics, the q_i denote coördinates and the p_i denote momenta.

Two functions $f(x)$ and $g(x)$ are said to be in involution, if the Poisson parenthesis,¹

$$(1) \quad (f, g) \equiv \sum_{i=1}^n \left(\frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} - \frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} \right) \equiv \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \frac{\partial g}{\partial x_{n+i}} - \frac{\partial f}{\partial x_{n+i}} \frac{\partial g}{\partial x_i} \right),$$

vanishes identically. On denoting by G the skew symmetric matrix, whose square is the identity,

$$G = \begin{pmatrix} 0 & +E \\ -E & 0 \end{pmatrix},$$

where E is the unit matrix of order n , (1) can be written in the more compact form

$$(2) \quad (f_x)' G g_x \equiv 0.$$

In (2) f_x and g_x denote the gradients of f and g respectively and $(f_x)'$ the transposed of the column vector f_x . A set of m forms f_1, f_2, \dots, f_m is called an *involutory system*, if any two of them are in involution, i. e., if $(f_i, f_j) \equiv 0$, $i, j = 1, 2, \dots, m$. One can readily verify that m is less than or equal to n , if the involutory system consists of independent functions,² independent in the

* Received March 18, 1940.

¹ Cf. e. g. E. T. Whittaker, *Analytical Dynamics* (Cambridge University Press (1904), page 288.

² Let $A'GA = 0$, where G is non-singular of order $2n$ and A is of rank r . Then, if $PAQ = B = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$, where E_r is the unit matrix of order r , $B'SB = 0$, where $G = P'SP$. Hence, if $S = (s_{ij})$, $i, j = 1, 2, \dots, 2n$, $s_{ij} = 0$, $i, j = 1, 2, \dots, r$. Since G and therefore S is non-singular, r is less than or equal to n .

ordinary function sense, and that the maximum value n of m may be obtained for suitably chosen independent functions.

If $h = h(x)$ is the Hamiltonian function of a conservative dynamical system, with the above notation we may write ³

$$(3) \quad G\dot{x} = h_x,$$

where $\dot{x} = dx/dt$.

If f is a conservative integral of (3), $(f_x)\dot{x}$ is identically zero in x or, since $G = -G^{-1}$, by (3)

$$(4) \quad (f_x)'Gh_x \equiv 0.$$

Conversely, if (4) is satisfied, $f = f(x)$ is a conservative integral of (3). Hence the m functions

$$f_1 = h, f_2, \dots, f_m$$

are m conservative integrals in involution, if, and only if,

$$(5) \quad (f_{i_x})'Gf_{j_x} \equiv 0, \quad (i, j = 1, 2, \dots, m).$$

It is known that $m = n$, but not more than n , conservative integrals in involution may be chosen to be independent in the functional sense mentioned above.

There remains the question: what becomes of these analytical facts in case the dynamical system is linear, i. e. if $h = h(x)$ is the quadratic form $\frac{1}{2}x'Hx$, where H is an arbitrary, but not zero, $2n$ -rowed symmetric matrix, representing the Hessian of h . In this case the Hamiltonian system appears in the simplified form

$$(6) \quad G\dot{x} = Hx.$$

Further the quadratic form $f = \frac{1}{2}x'Fx$ is by (4) an integral of (6), if, and only if,

$$(7) \quad x'FGHx \equiv 0.$$

Equation (7) is however equivalent to

$$FGH + H'G'F' = 0,$$

or, since F and H are symmetric and G skew symmetric, to

$$(8) \quad FGH = HGF.$$

Similarly the m quadratic forms, which belong to the symmetric matrices

³ Aurel Wintner, "On the linear conservative dynamical systems," *Annali di matematica pura ed applicata*, ser. 4, tomo 13 (1935-36), pp. 105-112.

$F_1 = H, F_2, \dots, F_m$ are m quadratic integrals of the system (6), forming an involutory system, if, and only if,⁴

$$(9) \quad F_i G F_j = F_j G F_i, \quad (i, j = 1, 2, \dots, m).$$

It is understood that all the matrices F_i are distinct from zero but are not necessarily non-singular.

By the general theorem, mentioned for non-linear systems, there always exist $m = n$ independent integrals in involution for the linear system (6). The conjecture was made by Professor Wintner that, in the case of a linear system, these $m = n$ integrals may be chosen to be quadratic forms. The main purpose of this paper is to show that this conjecture is correct—that for every $2n$ -rowed non-zero symmetric matrix H , there exist n symmetric matrices $F_1 = H, F_2, \dots, F_n$, which are independent and satisfy the involutory condition (9). It is understood that independence is now meant in the algebraic sense, i. e., that the corresponding quadratic forms are functionally independent.

By the general theory there always exist $2n - 1$ integrals, which are independent, and $n - 1$ may be obtained, by a theorem of Liouville from the n independent integrals in involution by means of quadratures and eliminations.⁵ In the linear case it is possible that some of these $n - 1$ integrals may also be quadratic; in fact, if the minimal equation of HG^{-1} is of degree $2m$, this number is $l = n - m$ and, if the degree of the minimal equation is $2m - 1$, the number is $l = n - m + 1$. The remaining $n - l - 1$ must then be determined by local elimination processes, which seem to lie outside the scope of an algebraic treatment.

It was found advisable first to determine the linearly independent quadratic integrals, a comparatively simple process; and then from them to determine the quadratic integrals independent in the more general sense. This was accomplished by the extensive use of linear differential operators, similar to the Aronhold operators of classical invariant theory.⁶ In § 6, when H is singular, the linear integrals of the system (6) are determined.

In the final section it is shown that the dynamical system, corresponding to the equations of variation of the small vibrations about an equilateral Lagrangian libration point in the restricted problem of three bodies, has, for all values of the masses, in addition to the energy integral only one quadratic integral; and this integral is determined.

The methods employed throughout the paper are purely algebraic, and the

⁴ Aurel Wintner, *loc. cit.*, page 108.

⁵ E. T. Whittaker, *op. cit.*, page 311.

⁶ E. g., L. E. Dickson, *Modern Algebraic Theories*, Chicago (1926), pp. 25-27.

proofs, to a large extent, are based on results, previously proved by the author,⁷ which give normal forms for a pencil of matrices, whose base consists of a symmetric and a non-singular skew symmetric matrix. These results, for convenience of reference, are given in § 1.

1. Normal forms. If H is a symmetric and G a non-singular skew-symmetric matrix, we shall say that the pair A, B is equivalent to the pair H, G , if there exists a non-singular matrix P , such that

$$PHP' = A \quad \text{and} \quad PGP' = B.$$

In normal form the matrices A and B of the pair equivalent to H, G are *similarly partitioned diagonal block*⁸ matrices

$$A = [A_1, A_2, \dots, A_k], \quad B = [B_1, B_2, \dots, B_k],$$

the blocks being determined by the elementary divisors of the matrix pencil $H - xG$. The elementary divisors of this pencil are subject to the following restrictions;⁹ if $(x - a)^r$, $a \neq 0$, occurs s times amongst the elementary divisors of the pencil, then so does the elementary divisor $(x + a)^r$ and the elementary divisor x^r , where r is odd, if it does occur, must occur an even number of times. Since the field of operations is the real field, there are four distinct forms for the matrices A_j and B_j .

Type (a). The pencil $A - xB$ has the single pair of real elementary divisors $(x - p)^r$, $(x + p)^r$. Then¹⁰

$$(10) \quad B_j = \begin{pmatrix} 0 & E_r \\ -E_r & 0 \end{pmatrix}, \quad A_j = \begin{pmatrix} L_j & 0 \\ 0 & -L_j' \end{pmatrix}, \quad B_j = \begin{pmatrix} 0 & L_j \\ L_j' & 0 \end{pmatrix},$$

where

$$(11) \quad L_j = pE_r + U_r.$$

In (11), E_r is the unit matrix of order r and U_r the auxiliary unit matrix of the same order. In particular, if r is odd, p may be zero.

Type (a₁). The pencil $A - xB$ has only the four elementary divisors $(\lambda \pm a \pm ib)^r$, $b \neq 0$. The matrices A_j and B_j are still determined by (10)

⁷ John Williamson, "On the algebraic problem concerning the normal forms of linear dynamical systems," *American Journal of Mathematics*, vol. 58 (January, 1936), pp. 141-163. The general field K is now the field R of all real numbers and the particular results now required are given in § 5, pp. 161-163.

⁸ The matrices A_j and B_j are square matrices of the same order.

⁹ John Williamson, *loc. cit.*, page 145 and page 162.

¹⁰ John Williamson, *loc. cit.*, page 158, formula (59).

and (11), if each unit is replaced by the two-rowed unit matrix, each zero by the two-rowed zero matrix and p by the matrix $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$.

Type (b). The pencil $A - xB$ has only the two pure imaginary divisors $(x - ib)^r$, $(x + ib)^r$. Then ¹¹

$$(12) \quad A_j = (pE_r + eU_r)B_j,$$

where e is the unit matrix of order 2, and

$$(13) \quad p = bi = \begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix},$$

$$(14) \quad B_j = i^r X_r,$$

$$(15) \quad X_r = \begin{pmatrix} 0 & 0 & . & 0 & 1 \\ 0 & 0 & . & -1 & 0 \\ . & . & . & 0 & 0 \\ (-1)^{r-1} & 0 & . & 0 & 0 \end{pmatrix}.$$

Type (b₁). The pencil $A_j - xB_j$ has the single elementary divisor x^{2r} . Then $A_j = U_{2r}$ and $B_j = X_{2r}$.

For later purposes we require the following. If r is even, X_r is skew symmetric and, if r is odd, X_r is symmetric and therefore for all values of r the matrix B_j in (14) is skew symmetric. Further

$$(16) \quad X_r U_r = -U'_r X_r$$

and, since $X_r^2 = \pm E_r$,

$$(17) \quad U_r X_r = -X_r U'_r.$$

In *type (a)*, when $p \neq 0$, any matrix D commutative with $A_j B_j^{-1}$ is of the form $\begin{pmatrix} D_{11} & 0 \\ 0 & D_{22} \end{pmatrix}$, where

$$(18) \quad D_{11} = \sum_{k=0}^{r-1} f_k U_r^k \quad \text{and} \quad D_{22} = \sum_{k=0}^{r-1} g_k U'_r{}^k.$$

If $p = 0$, the matrices defined by (18) are certainly commutative with $A_j B_j^{-1}$ but of course are not the only ones.

In *type (a₁)* D has the same form except that f_k and g_k are both poly-

¹¹ John Williamson, *loc. cit.*, page 155, formula (55). Formula (55) is of course simplified for this special case as indicated on page 162. The fact that B_j is not unique but may be replaced by $-B_j$ does not alter the form of a matrix commutative with B_j .

nomials in p , i. e., are of the form $\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$. In type (b), D has the form D_{11} in (18), where f_k is again a polynomial¹² in p .

2. We now consider the purely algebraic problem of determining the number m of linearly independent symmetric matrices F_i of order $2n$, which satisfy

$$(19) \quad F_i G H = H G F_i, \quad (i = 1, 2, \dots, m),$$

where H is a given symmetric matrix and $G = \begin{pmatrix} 0 & +E \\ -E & 0 \end{pmatrix}$ is the non-singular skew symmetric matrix mentioned in the introduction. If (19) is satisfied,

$$F_i G H G = H G F_i G,$$

or, since $G = -G^{-1}$,

$$(20) \quad F_i G^{-1} H G^{-1} = H G^{-1} F_i G^{-1}.$$

Hence, if F_i satisfies (19), $F_i G^{-1}$ is commutative with $H G^{-1}$. Since the number and nature of the linearly independent matrices $F_i G^{-1}$, commutative with $H G^{-1}$, are known,¹³ it is only necessary to determine for which of these known matrices $F_i G^{-1}$ the matrix F_i is symmetric.

The number of linearly independent matrices commutative with $H G^{-1}$ depends on the number and the nature of the elementary divisors of the matrix pencil $H - xG$. Hence, in considering the general case, it is necessary to reduce H and G to the normal forms given in section 1. However, if $H G^{-1}$ is not derogatory, i. e., if the minimal equation of $H G^{-1}$ is the same as its characteristic equation, any matrix commutative with $H G^{-1}$ is a polynomial¹⁴ in $H G^{-1}$. A maximal set of linearly independent matrices commutative with $H G^{-1}$ therefore contains exactly $2n$ members; and one such set consists of the $2n$ distinct powers of $H G^{-1}$, i. e., of the $2n$ matrices

$$(H G^{-1})^k, \quad (k = 0, 1, 2, \dots, 2n - 1).$$

We may therefore suppose that

$$F_i G^{-1} = (H G^{-1})^i$$

¹² J. H. M. Wedderburn, "Lectures on matrices," *American Mathematical Society Colloquium Publications*, vol. 17 (1934), page 124; John Williamson, "The idempotent and nilpotent elements of a matrix," *American Journal of Mathematics*, vol. 58 (1936), p. 477.

¹³ J. H. M. Wedderburn, *op. cit.*, page 105.

¹⁴ J. H. M. Wedderburn, *op. cit.*, page 27; C. C. MacDuffee, *The Theory of Matrices*, Berlin (1933), page 94.

or that

$$F_i = (HG^{-1})^{i-1}H.$$

Consequently, F'_i the transposed of F_i satisfies

$$F'_i = (-1)^{i-1}H(G^{-1}H)^{i-1} = (-1)^{i-1}(HG^{-1})^{i-1}H = (-1)^{i-1}F_i.$$

If F_i is symmetric, $i-1$ must be even and i must be odd. Therefore, if HG^{-1} is not derogatory, there exist exactly n linearly independent symmetric matrices F_i which satisfy (19). One such set consists of the n matrices¹⁵

$$(21) \quad F_i = (HG^{-1})^{2(i-1)}H, \quad (i = 1, 2, \dots, n).$$

Since the matrices F_iG^{-1} , where F_i is defined in (21), are all polynomials in HG^{-1} , it follows that

$$F_iGF_j = F_jGF_i, \quad (i, j = 1, 2, \dots, n),$$

and hence, that the n quadratic integrals corresponding to the matrices F_i form a set of integrals in involution. Consequently, we have

THEOREM I. *If HG^{-1} is not derogatory, there exist n linearly independent quadratic integrals of the system (6). These n quadratic integrals form a set in involution and may be so chosen that the corresponding matrices are the matrices F_i in (21).*

It will be shown later (§ 3) that these n quadratic integrals are not only linearly independent but also functionally independent.

If a matrix F , which satisfies (8), is not symmetric, we find, on taking the transposed of both sides of (8), that

$$F'G'H' = H'G'F',$$

or, since H is symmetric and G skew symmetric, that

$$F'GH = HGF'.$$

Therefore we have

LEMMA 1. *If a matrix F satisfies (8), so does the matrix F' , the transposed of F .*

If the pencil of matrices $H - xG$ is congruent to the pencil $A - xB$, so that there exists a non-singular matrix P satisfying both of the equations,

$$PHP' = A \quad \text{and} \quad PGP' = B,$$

¹⁵ Since $G^{-1} = -G$ the matrix $F_i = (HG)^{2(i-1)}H$. The matrix G^{-1} is used instead of $-G$ to emphasize the fact, that it is the matrix pencil $H - xG$ which is the dominating factor throughout this discussion.

then

$$AB^{-1}Q_iB^{-1} = Q_iB^{-1}AB^{-1},$$

where $Q_i = PF_iP'$.

Consequently, we may replace the matrices H and G in (19) or (20) by any pair A, B equivalent to H, G or, what is equivalent to this, we may suppose that the pair H, G is already in the normal form described in § 1. We therefore do this and assume that H and G are the matrices A and B respectively of § 1. We let

$$F = (F_{ij}), \quad (i, j = 1, 2, \dots, k),$$

be a partition of F similar to that of A or B and therefore, as a consequence of (20), have

$$(22) \quad F_{ij}B_j^{-1}A_jB_j^{-1} = A_iB_i^{-1}F_{ij}B_j^{-1}, \quad (i, j = 1, 2, \dots, k).$$

If FG^{-1} is the most general matrix commutative with HG^{-1} , by Lemma 1, we obtain the most general symmetric matrix satisfying (8) from this by putting $F_{ji} = F'_{ij}$, $i < j$, and restricting F_{ii} to be symmetric. If $A_iB_i^{-1}$ is non-derogatory, it is a consequence of Theorem 1, that the number of linearly independent matrices $F_{ii}B_i^{-1}$ commutative with $A_iB_i^{-1}$, for which F_{ii} is symmetric, is exactly one half the order of A_i , i.e. is one half the number of linearly independent matrices commutative with $A_iB_i^{-1}$. Consequently, if each of the matrices $A_iB_i^{-1}$ is non-derogatory, the number of linearly independent symmetric matrices F , which satisfy (8), is exactly one half the total number of linearly independent matrices commutative with HG^{-1} and as remarked earlier, this number is known.¹⁶ A maximal set of symmetric matrices F_i , which satisfy (9), must consist entirely of diagonal block matrices

$$[F_{11}, F_{22}, \dots, F_{kk}];$$

and, as a consequence of Theorem 1, the number of linearly independent matrices in such a set is n . It is apparent from § 1, that $A_jB_j^{-1}$ is derogatory, if, and only if, the pencil $A_j - xB_j$ is of type (a) with $p = 0$, and that then H is singular. Accordingly, we have proved

THEOREM 2. *If H is non-singular, there exist n linearly independent quadratic integrals of the system (6), which form a set in involution. The number of linearly independent quadratic integrals of the system (6) is exactly one half the number of linearly independent matrices commutative with HG^{-1} .*

If $A_jB_j^{-1}$ is derogatory, the elementary divisors of the pencil $A_j - xB_j$

¹⁶ J. H. M. Wedderburn, *op. cit.*, page 105.

are x^r, x^r where r is odd. On dropping the suffix r , we obtain from equation (10)

$$(23) \quad A_j B_j^{-1} = \begin{pmatrix} U & 0 \\ 0 & -U \end{pmatrix}.$$

If

$$F_{jj} = D B_j^{-1},$$

as a consequence of (22), we have

$$(24) \quad D[U, -U'] = [U, -U']D.$$

Finally, if $D = (D_{ij})$, $i, j = 1, 2$, is a partition of D similar to that of $A_j B_j^{-1}$ in (23), then (24) yields the equations

$$(25) \quad D_{11}U = UD_{11}, \quad D_{22}U' = U'D_{22}, \quad -D_{12}U' = UD_{12}, \quad -D_{21}U = U'D_{21}.$$

The matrix F_{jj} therefore has the form

$$\begin{pmatrix} D_{21} & D_{22} \\ -D_{11} & D_{12} \end{pmatrix}$$

and is symmetric, if, and only if, D_{21} and D_{12} are symmetric and $D_{11} = -D'_{22}$. By (16), $U' = -XUX^{-1}$ and therefore by (25), $D_{12}XUX^{-1} = UD_{12}$; so that $D_{12}X$ is commutative with U and is therefore a polynomial in U . Hence

$$D_{12} = \sum_{i=0}^{r-1} f_i U^i X^{-1}.$$

Since r is odd, X is symmetric and therefore,

$$D'_{12} = (X^{-1}) \sum_{i=0}^{r-1} f_i U' = \sum_{i=0}^{r-1} (-1)^i f_i U^i X^{-1}.$$

If D_{12} is symmetric, $f_i = 0$ when i is odd and therefore, if $r = 2m + 1$, D_{12} depends on the $m + 1$ parameters f_i , $i = 0, 2, 4, \dots, 2m$. Similarly, if D_{21} is symmetric, D_{21} depends on $m + 1$ parameters. Finally, if D_{11} is the most general matrix commutative with U , then D_{11} depends on $r = 2m + 1$ parameters and D'_{11} is the most general matrix commutative with U' . Hence, if $D'_{11} = -D_{22}$, the matrix pair D_{11} and D_{22} depends on only $2m + 1$ parameters. Therefore the matrix F_{jj} , when it is symmetric, depends on

$$m + 1 + m + 1 + 2m + 1 = 4m + 3$$

parameters. The general matrix $F_{jj} B_j^{-1}$, commutative with $A_j B_j^{-1}$, however, depends on $4r = 8m + 4$ parameters and $4m + 3 = \frac{1}{2}(4r) + 1$. For example, in the simplest case, $r = 1$, A_j is the zero matrix and F_{jj} is of course arbitrary. To restrict F_{jj} to be symmetric imposes only one condition and there are,

therefore, in this simple case $3 = \frac{1}{2}4 + 1$ linearly independent symmetric matrices F_{jj} . Consequently, if $A_j B_j^{-1}$ is derogatory, the number of linearly independent symmetric matrices F_{jj} is one more than half the number of linearly independent matrices commutative with $A_j B_j^{-1}$. Moreover, the matrices F_{jj} , for which $D_{12} = 0$, form a set in involution, since any two such matrices $F_{jj} B_j^{-1}$ are obviously commutative. Their number is $r = 2m + 1$ and we therefore have

THEOREM 3. *The number of linearly independent symmetric matrices F_i , which satisfy (19), is exactly one half the number of linearly independent matrices commutative with HG^{-1} , unless the pencil $H - xG$ has a pair of elementary divisors of the form (x^{2m+1}, x^{2m+1}) . For each such pair of elementary divisors, the number of linearly independent symmetric matrices F_i is increased by one.*

It is obvious that two matrices FG^{-1} for which $F_{ij} = 0, i \neq j$, are always commutative and it is known that a maximal set of matrices FG^{-1} , commutative in pairs, consists solely of matrices¹⁷ for which $F_{ij} = 0, i \neq j$. The number of linearly independent symmetric matrices F in such a set is, therefore, $d = \sum_{i=1}^k d_i$, where d_i is the number of linearly independent matrices F_{ii} . Since we have just shown that $d_i = \frac{1}{2}n_i$, where n_i is the order of F_{ii} , $d = \sum_{i=1}^k \frac{1}{2}n_i = \frac{1}{2}2n = n$. We have accordingly proved

THEOREM 4. *Every linear conservative dynamical system with n degrees of freedom has at least n linearly independent quadratic integrals in involution.*

3. While the results obtained up to the present all deal with linear independence, we now determine, from the known linearly independent quadratic integrals or forms, all the functionally independent quadratic integrals. We first show that n quadratic integrals in involution, which are linearly independent, are necessarily functionally independent.

We note that, if

$$F = [F_{11}, F_{22}, \dots, F_{kk}]$$

is a diagonal block symmetric matrix, the quadratic forms corresponding to the matrices

$$[F_{11}, 0, 0, \dots, 0], [0, F_{22}, \dots, 0], \dots, [0, 0, \dots, F_{kk}]$$

are not only linearly, but also functionally independent, since each involves

¹⁷ J. H. M. Wedderburn, *op. cit.*, page 106.

a different set of variables. Accordingly, we need only consider four simple cases—those corresponding to the types (a) , (a_1) , (b) and (b_1) of section 1.

Type (a). Let $H - xG$ have the single pair of real elementary divisors $(\lambda \pm p)^n$. Then we may take H in the normal form given by (10),

$$H = \begin{pmatrix} L & 0 \\ 0 & -L' \end{pmatrix} G,$$

where L is the matrix L_i of (11). Then the n linearly independent matrices F of Theorem 4 may be chosen to be¹⁸

$$(26) \quad \begin{pmatrix} 0 & U^k \\ U'^k & 0 \end{pmatrix}, \quad (k = 1, 2, \dots, n),$$

and those of Theorem 1 as

$$(27) \quad \begin{pmatrix} 0 & L^k \\ L'^k & 0 \end{pmatrix}, \quad (k = 0, 1, 2, \dots, n-1).$$

The quadratic forms corresponding to the symmetric matrices (26) are

$$(28) \quad 2f_j = 2 \sum_{p=1}^j x_p x_{2n-j+p}, \quad (j = n-k = 1, 2, \dots, n).$$

The n quadratic forms (28) obviously are functionally independent, since each of the forms f_1, f_2, \dots, f_n contains a variable, which does not occur in any of its predecessors. Since the matrices (26) are all linear combinations of the matrices (27), the n quadratic forms corresponding to the matrices (27) are also functionally independent. It should be noted that the above results are true, even if $p = 0$; in this case, however, there do exist in addition other symmetric matrices F , which satisfy (8) but are not of the above form.

Type (a₁). Let $H - xG$ have only the four elementary divisors $(x \pm a \pm ib)^n$, $b \neq 0$, so that H is now of order $4n$. We may take H and G in the normal form described in section 1. Then the $2n$ linearly independent matrices F of Theorem 1 are given by (27), with $k = 0, 1, 2, \dots, 2n-1$, while the matrices of Theorem 4 are the matrices (26) together with the matrices

$$(29) \quad \begin{pmatrix} 0 & iU^k \\ -iU'^k & 0 \end{pmatrix}, \quad i = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

For convenience we relabel the $4n$ variables x in the order $x_1, \xi_1, x_2, \xi_2, \dots, x_{2n}, \xi_{2n}$. Then the $2n$ linearly independent quadratic forms corresponding to the matrices (26) and (29) are respectively

¹⁸ J. H. M. Wedderburn, *op. cit.*, page 104.

$$(30) \quad 2f_j = 2 \sum_{p=1}^j (x_p x_{2n-j+p} + \xi_p \xi_{2n-j+p}), \quad (j = 1, 2, \dots, n),$$

and

$$(31) \quad 2g_j = 2 \sum_{p=1}^j (x_p \xi_{2n-j+p} - \xi_p x_{2n-j+p}).$$

On arranging these $2n$ quadratic forms in the order,

$$f_1, g_1, f_2, g_2, \dots, f_n, g_n,$$

we see that they are functionally independent; for f_k and g_k are functionally independent and each pair f_k, g_k contains two variables which do not occur in any of the preceding pairs.

Type (b₁). Let the pencil $H - xG$ have the single pair of elementary divisors $(x \pm ib)^n$, $n \neq 0$. Then we may take H and G in the normal forms given by (12), (13) and (14), (15) respectively. On dropping the suffix r , we can easily show that any matrix F satisfying (8) is of the form $D_{11}X$, where D_{11} is given by (18), and the f_k are two-rowed matrices of the form $\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$. If F is to be symmetric, $d = 0$ when $n - k$ is odd, while $c = 0$ when $n - k$ is even. If the $2n$ variables x are relabelled in the order, $x_1, \xi_1, x_2, \xi_2, \dots, x_n, \xi_n$, the n linearly independent quadratic forms of Theorem 4 may be taken as those n of the $2n$ quadratic forms,

$$(32) \quad f_j = \sum_{p=1}^j (-1)^{p+1} (x_p x_{j+1-p} + \xi_p \xi_{j+1-p}), \quad (j = 1, 2, \dots, n),$$

and

$$(33) \quad g_j = \sum_{p=1}^j (-1)^{p+1} (x_p \xi_{j+1-p} - \xi_p x_{j+1-p}), \quad (j = 1, 2, \dots, n),$$

which do not vanish identically. In (32), f_j is zero, if j is even; similarly, in (33), g_j is zero, if j is odd. If we write the forms in the order $f_1, g_2, f_3, g_4, \dots$ it follows, as in the previous cases, that those n of the forms (32) and (33), which are not identically zero, are functionally independent.

Type (b₂). Let $H - xG$ have the single elementary divisor x^{2n} . Then, with the notation of § 1 *type (b₂)* we may take $G = X$ and $H = U$. The matrices of the n linearly independent quadratic forms are then $U^{2i}X^{-1}$, $i = 0, 1, 2, \dots, n-1$, by Theorem 1. The corresponding quadratic forms are proportional to

$$(34) \quad f_j = \sum_{p=1}^j (-1)^{p+1} x_p x_{j+1-p}, \quad (j = 1, 3, 5, \dots, 2n-1).$$

As in previous cases, the n quadratic forms (34) are functionally independent. Combining the separate results of this section we have

THEOREM 5. *Every linear conservative dynamical system with n degrees of freedom has at least n functionally independent quadratic integrals in involution. If HG^{-1} is not derogatory, the quadratic forms, whose matrices are $(HG^{-1})^{2(j-1)}H$, $j = 1, 2, \dots, n$, are n functionally independent quadratic integrals in involution of the linear system (6).*

4. In determining the maximal number of functionally independent quadratic integrals (not necessarily in involution), we shall once again start from the known set of linearly independent ones. As in the previous section,¹⁹ it will only be necessary to consider four special cases:

Type (a). Let $H - xG$ have the elementary divisors

$$(x - p)^{e_1}(x + p)^{e_1}; \quad (i = 1, 2, \dots, k; e_1 \geq e_2 \geq \dots \geq e_k);$$

p real and different from zero. Then we may take H and G in the normal form of § 1, where A_j and B_j are given by (10) and (11) for all values of j . Therefore, if F is a symmetric matrix which satisfies (8),

$$F = \begin{pmatrix} 0 & W \\ W' & 0 \end{pmatrix}, \text{ where } W = (W_{ij}), \quad (i, j = 1, 2, \dots, k)$$

and

$$(35) \quad W_{ij} = \begin{pmatrix} G_{ij} \\ 0 \end{pmatrix}, \quad e_i \geq e_j; \quad W_{ij} = (0, G_{ij}), \quad e_i \leq e_j.$$

The matrix G_{ij} is of the form²⁰

$$(36) \quad G_{ij} = \sum_{s=0}^{e-1} g_{ijs} U^s,$$

where e is the minimum of e_i and e_j , and U is the auxiliary unit matrix of order e . If all g_{abc} are zero except a particular one, g_{ijs} , which has the value unity, we denote the corresponding matrix F by $F_{ij e-s}$ and the corresponding quadratic form by $2f_{ij e-s}$. With this notation the linearly independent quadratic forms are

$$(37) \quad f_{ijt} = \sum_{p=1}^t x_{\tau+p} x_{\sigma+t-p}, \quad (i, j = 1, 2, \dots, k; t = 1, 2, \dots, e-1),$$

where $\tau = e_1 + e_2 + \dots + e_{i-1}$ and $\sigma = e_1 + e_2 + \dots + e_j$. The quadratic

¹⁹ If $AB = BA$ and A is the diagonal block matrix $[A_1, A_2]$ where the minimal equations of A_1 and A_2 are relatively prime, B is also a diagonal block matrix partitioned similarly to A .

²⁰ J. H. M. Wedderburn, *op. cit.*, page 104.

forms (37) are actually of the same nature as those of (28), except that the sets of variables involved are no longer x_1, x_2, \dots, x_n and x_{2n}, \dots, x_{n+1} . In particular, if $x_{\tau+1} = y_i$ and $x_{n+\sigma} = z_j$,

$$(38) \quad f_{ij1} = x_{\tau+1}x_{n+\sigma} = y_i z_j.$$

In order to determine the functionally independent quadratic integrals, we make use of linear differential operators reminiscent of the Aronhold operators of classical invariant theory.²¹ We define the linear differential operators Ω_i and Ω_j by

$$(39) \quad \Omega_i = \sum_{p=1}^{e_i} p x_{\tau+p+1} \frac{\partial}{\partial x_{\tau+p}} + \sum_{p=0}^{e_i-1} (e_i - p) x_{n+\tau+p} \frac{\partial}{\partial x_{n+\tau+p+1}},$$

$\tau = e_1 + e_2 + \dots + e_{i-1}$, and

$$\Omega_j = \sum_{p=1}^{e_j} p x_{\rho+p+1} \frac{\partial}{\partial x_{\rho+p}} + \sum_{p=0}^{e_j-1} (e_j - p) x_{n+\rho+p} \frac{\partial}{\partial x_{n+\rho+p+1}},$$

$\rho = e_1 + e_2 + \dots + e_{j-1}$. Then

$$\Omega_i f_{ijt} = \sum_{p=1}^t p x_{\tau+p+1} x_{n+\sigma-t+p},$$

and

$$\Omega_j f_{ijt} = \sum_{p=1}^t (t - p + 1) x_{\tau+p} x_{n+\sigma-t-1+p},$$

since $\sigma = \rho + e_j$. Therefore

$$\begin{aligned} (\Omega_i + \Omega_j) f_{ijt} &= \sum_{p=2}^{t+1} (p-1) x_{\tau+p} x_{n+\sigma-t-1+p} + \sum_{p=1}^t (t-p+1) x_{\tau+p} x_{n+\sigma-t-1+p} \\ &= t \sum_{p=1}^{t+1} x_{\tau+p} x_{n+\sigma-t-1+p}, \end{aligned}$$

and finally by (37)

$$(40) \quad (\Omega_i + \Omega_j) f_{ijt} = t f_{ijt+1}.$$

We note that the $2n - e_1$ quadratic forms

$$(41) \quad f_{ijt_j}; \quad (t_j = 1, 2, \dots, e_j; j = i \text{ or } i+1; i = 1, 2, \dots, k),$$

are functionally independent; for, if we arrange them in the order

$$f_{111} f_{112}, \dots, f_{11e_1}; f_{121}, \dots, f_{12e_2}; f_{221}, \dots, f_{22e_2}; \dots, f_{kke_k},$$

each form contains a variable which does not occur in any of its predecessors. These variables are in order

$$x_1, x_2, \dots, x_{e_1}; x_{n+e_1+e_2}, \dots, x_{n+e_1+1}; x_{e_1+1}, \dots, x_{e_1+e_2}; \dots, x_n.$$

²¹ Cf. L. E. Dickson, *Modern Algebraic Theories*, pp. 25-27.

Next we show that every quadratic form (37) is a function of the quadratic forms (41). It is a consequence of (38) that $z_j/z_{j+1} = f_{jj1}/f_{jj+11}$ and therefore that

$$(42) \quad z_j = q_{ij} z_i,$$

where q_{ij} is a rational function of the quadratic forms f_{ss1} and f_{ss+11} where

$$(43) \quad \text{either } j \leq s \leq i-1 \text{ or } i \leq s \leq j-1.$$

Consequently we have

$$(44) \quad f_{ij1} = q_{ij} f_{ii1}$$

and are in a position to prove,

LEMMA 2. *The quadratic form f_{ijt} is a rational function of the quadratic forms f_{ssr} , f_{ss+1r} , where $r \leq t$, and s satisfies one of the inequalities (43).*

We shall prove this lemma by induction on t and observe that, as a consequence of (44), it is true when $t = 1$. We assume the lemma true for the value t and therefore have

$$(45) \quad f_{ijt} = R = R(f_{ssr}; f_{s,s+1,r}), \quad r \leq t, \quad s \text{ defined by (43)}.$$

Let $\Omega = \Sigma \Omega_q$ where Ω_q is defined in a similar manner to Ω_i in (39) and the summation extends from i to j or from j to i . Since, by definition, Ω is a linear operator, ΩR is a sum of terms, each term being a product of the partial derivative of R with respect to one of its variables f_{spr} by Ωf_{spr} . By (40)

$$\Omega f_{ijt} = t f_{ijt+1} \quad \text{and} \quad \Omega f_{spr} = r f_{spr+1}$$

and therefore by operating with Ω on both sides of equation (45) we have $t f_{ijt+1} = W$, where W is a rational function of f_{ssr} and f_{ss+1r} ; s is defined by (43) and $r \leq t+1$. Hence our lemma is proved and consequently all the quadratic forms (37) are functions of the $2n - e_1$ functionally independent quadratic forms (41). It should be noted, for later reference, that e_1 is the highest exponent of $(x \pm a)$ in the elementary divisors of the pencil $H - xG$ or of $HG^{-1} - xE$, and that accordingly the minimal equation of HG^{-1} is of degree $2e_1$.

Type (a_1) . Let $H - xG$ have the elementary divisors

$$(x \pm a \pm ib)^{e_i}; \quad (i = 1, 2, \dots, k; \quad b \neq 0; \quad e_1 \geq e_2 \geq \dots \geq e_k).$$

Then, if we let G be of order $4n$ and relabel the variables $x_1, \xi_1, x_2, \xi_2, \dots, x_{2n}, \xi_{2n}$ we find, by an argument similar to that applied to the forms (30) and (31) of the previous section, that the forms

$$(46) \quad w_{ijt} = \sum_{p=1}^t (x_{\tau+p} x_{n+\sigma-t+p} + \xi_{\tau+p} \xi_{n+\sigma-t+p}),$$

and

$$(47) \quad v_{ijt} = \sum_{p=1}^t (x_{\tau+p} \xi_{n+\sigma-t+p} - \xi_{\tau+p} x_{n+\sigma-t+p})$$

are linearly independent if t , σ and τ have the values given in (37). In particular we may write

$$(48) \quad w_{ij1} = y_i z_j + \eta_i \xi_j \quad \text{and} \quad v_{ij1} = y_i \xi_j - \eta_i z_j.$$

The $4n - 2e_1$ quadratic forms (46) and (47), for which $j = i$ or $i + 1$, are functionally independent, for they can be so arranged in pairs that each pair contains two variables which do not occur in any of the preceding pairs.

On forming the Jacobian of the eight forms

$$(49) \quad w_{ik1}, -v_{ik1}, w_{ij1}, -v_{ij1}, w_{jk1}, -v_{jk1}, w_{jj1}, -v_{jj1},$$

we obtain, with the notation of (48), the eight-rowed square matrix

$$K = \begin{pmatrix} Z_k & 0 & Y_i & 0 \\ Z_j & 0 & 0 & Y_i \\ 0 & Z_k & Y_j & 0 \\ 0 & Z_j & 0 & Y_j \end{pmatrix}, \quad \text{where } Z_k = \begin{pmatrix} z_k & \xi_k \\ -\xi_k & z_k \end{pmatrix} \quad \text{and } Y_i = \begin{pmatrix} y_i & \eta_i \\ \eta_i & -y_i \end{pmatrix}.$$

$$\text{If} \quad W_i = Y_i \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} y_i & -\eta_i \\ \eta_i & y_i \end{pmatrix},$$

the rank of K is the same as that of the matrix P obtained from K by replacing Y_i by W_i and Y_j by W_j . Since the matrices W_i and Z_j are commutative, a simple calculation shows that the product of the two by four matrix

$$(-Y_j Z_j, Y_j Z_k, +Y_i Z_j, -Y_i Z_k)$$

by P is the zero matrix. Hence, at most six of the eight forms (49) are functionally independent and, since any three of the pairs w_{ab1}, v_{ab1} in (49) are functionally independent, each member of the fourth pair is a function of the other six forms. In particular

$$(50) \quad w_{ik1} = F(w_{ij1}, v_{ij1}; w_{jk1}, v_{jk1}, w_{jj1}, v_{jj1}),$$

and, if $k = i$,

$$(51) \quad w_{ji1} = G(w_{ij,1}, v_{ij,1}; w_{ii,1}, v_{ii,1}; w_{jj1}, v_{jj1})$$

with similar results for the corresponding forms v_{ik1} and v_{ji1} .

As a consequence of (51), if $a > b$, w_{ab1} is a function of forms w_{ij1} and v_{ij1} where $i \leq j$. As a consequence of (50), when $b - a > 2$, w_{ab1} is a

function of forms w_{ij1} and v_{ij1} where $j-i < b-a$. The same is true of the forms v_{ab1} and therefore we have

$$(52) \quad w_{ij1} = F(w_{ss1}, v_{ss1}; w_{s,s+1,1}, v_{s,s+1,1}), \quad v_{ij1} = G(w_{ss1}, v_{ss1}; w_{s,s+1,1}, v_{s,s+1,1}).$$

Let O_i and O_j be the differential operators obtained from Ω_i and Ω_j , defined in (39), by replacing x by ξ . Then

$$(53) \quad (\Omega_i + O_i + \Omega_j + O_j)w_{ijt} = tw_{ijt+1},$$

and

$$(54) \quad (\Omega_i + O_i + \Omega_j + O_j)v_{ijt} = tv_{ijt+1}.$$

If $O = \sum O_p$, where the summation extends from i to j or j to i , we may prove a lemma, which is the analogue of Lemma 2. In the latter the operator Ω is replaced by $\Omega + O$, (44) by (52), and (40) by (53) or by (54). It then follows that any of the quadratic forms (46) or (47) is a function of the $4n - 2e_1$ quadratic forms for which $j = i$ or $i + 1$. Therefore, in this particular case, where G is of order $4n$ and the minimal equation of HG^{-1} is of degree $4e_1$, the number of functionally independent quadratic integrals of the system (6) is $4n - 2e_1$.

Type (b₁). Let the pencil $H - xG$ have the elementary divisors

$$(x \pm ib)^{e_i}, \quad (b \neq 0, j = 1, 2, \dots, k; e_1 \geq e_2 \geq \dots \geq e_k).$$

Then we may take H and G in the normal form of § 1, where A_j and B_j are given by equations (12), (13), and (14). If F is a symmetric matrix, which satisfies (8), and, if $F = (F_{ij})$, $i, j = 1, 2, \dots, k$ is a partition of F similar to that of H or G ,

$$F_{ij} = (W_{ij})X_j,$$

where W_{ij} is defined by (35) and (36), with the addition that

$$g_{ijs} = \begin{pmatrix} r_{ijs} & t_{ijs} \\ -t_{ijs} & r_{ijs} \end{pmatrix}.$$

Since F is symmetric, $F'_{ji} = F_{ij}$, so that we need only concern ourselves with the case in which $i \leq j$. If $F_{ab} = 0$ for all a and b except when $a = i$, $b = j$ and $a = j$, $i = b$, the corresponding quadratic form involves two sets of variables, one containing $2e_i$, and the other $2e_j$ variables. If for convenience of notation we write $e_i = e$ and $e_j = d$, we can denote these sets by

$$x_1, \xi_1, x_2, \xi_2, \dots, x_e, \xi_e \quad \text{and} \quad y_1, \eta_1, y_2, \eta_2, \dots, y_d, \eta_d$$

respectively. The corresponding linearly independent quadratic forms are then

$$(55) \quad w_{ijt} = \sum_{p=1}^t (-1)^{p+1} (x_p y_{t+1-p} + \xi_p \eta_{t+1-p}), \quad (t = 1, 2, \dots, d),$$

and

$$(56) \quad v_{ijt} = \sum_{p=1}^t (-1)^{p+1} (x_p \eta_{t+1-p} - \xi_p y_{t+1-p}), \quad (t = 1, 2, \dots, d).$$

If $i = j$, the variables x , ξ are the same as y , η ; and $w_{ijt} = 0$, if t is even, while $v_{iit} = 0$, when t is odd. Let

$$\Omega_x = \sum_{p=1}^e p \left(\xi_{p+1} \frac{\partial}{\partial x_p} - x_{p+1} \frac{\partial}{\partial \xi_p} \right) \quad \text{and} \quad \Omega_y = \sum_{p=1}^d p \left(\eta_{p+1} \frac{\partial}{\partial y_p} - y_{p+1} \frac{\partial}{\partial \eta_p} \right).$$

Then

$$\begin{aligned} \Omega_x w_{ijt} &= \sum_{p=1}^t (-1)^{p+1} p (\xi_{p+1} y_{t+1-p} - x_{p+1} \eta_{t+1-p}) \\ &= \sum_{p=2}^{t+1} (-1)^p (p-1) (\xi_p y_{t+2-p} - x_p \eta_{t+2-p}), \end{aligned}$$

while

$$\Omega_y w_{ijt} = \sum_{p=1}^t (-1)^{p+1} (t+1-p) (x_p \eta_{t+2-p} - \xi_p y_{t+2-p}).$$

Therefore

$$(57) \quad (\Omega_x + \Omega_y) w_{ijt} = t \sum_{p=1}^{t+1} (-1)^{p+1} (x_p \eta_{t+2-p} - \xi_p y_{t+2-p}) = t v_{ijt+1}.$$

Similarly

$$\Omega_x v_{ijt} = \sum_{p=1}^t (-1)^{p+1} p (\xi_{p+1} \eta_{t+1-p} + x_{p+1} y_{t+1-p}),$$

and

$$\Omega_y v_{ijt} = \sum_{p=1}^t (-1)^{p+1} (t+1-p) (-x_p y_{t+2-p} - \xi_p \eta_{t+2-p}),$$

so that

$$(58) \quad (\Omega_x + \Omega_y) v_{ijt} = -t w_{ijt+1}.$$

The Jacobian of the four forms w_{i11} , w_{j11} , w_{i1j} and v_{i1j} is the matrix

$$\begin{pmatrix} x_1 & \xi_1 & 0 & 0 \\ 0 & 0 & y_1 & \eta_1 \\ y_1 & \eta_1 & x_1 & \xi_1 \\ \eta_1 & -y_1 & -\xi_1 & x_1 \end{pmatrix}.$$

The determinant of this matrix is zero while the matrix of its first three rows has rank three. Therefore v_{i1j} is a function of w_{i11} , w_{j11} and w_{i1j} . Since $\Omega_x + \Omega_y$ is a linear operator, we may employ an argument similar to that used in the proof of Lemma 2, and from (57) and (58) deduce that w_{ij2} is a function of w_{i11} , w_{j11} , w_{i1j} , v_{i12} , v_{j12} , v_{i1j} . By repeating this argument we finally have the result that, for any value of k , w_{ij2k} and v_{ij2k+1} are both functions of forms of the type f_{abc} , where

$$(59) \quad f_{abc} = w_{abc}, \text{ if } c \text{ is odd, and } f_{abc} = v_{abc}, \text{ if } c \text{ is even.}$$

If we denote the variables associated with $F_{\tau\tau}$ by z and ξ , the Jacobian matrix of the six forms

$$(60) \quad f_{i\tau 1}, f_{jj 1}, f_{\tau\tau 1}, f_{ij 1}, f_{j\tau 1}, f_{i\tau 1},$$

formed with respect to the variables $x_1, y_1, z_1, \xi_1, \eta_1, \zeta_1$, is the matrix

$$Q = \begin{pmatrix} x_1 & 0 & 0 & \xi_1 & 0 & 0 \\ 0 & y_1 & 0 & 0 & \eta_1 & 0 \\ 0 & 0 & z_1 & 0 & 0 & \zeta_1 \\ y_1 & x_1 & 0 & \eta_1 & \xi_1 & 0 \\ 0 & z_1 & y_1 & 0 & \zeta_1 & \eta_1 \\ z_1 & 0 & x_1 & \zeta_1 & 0 & \xi_1 \end{pmatrix}.$$

If X is the column vector with components $\xi_1, \eta_1, \zeta_1, -x_1, -y_1, -z_1$, the vector QX is the zero vector. Hence Q is singular. The first five of the forms (60) are functionally independent and therefore $f_{i\tau 1}$ is a function of these five. Therefore, if $i+1 < \tau$,

$$(61) \quad f_{i\tau 1} = F(f_{ab1}),$$

where $b-a < \tau-1$.

As a consequence of (61) we have

$$(62) \quad f_{i\tau 1} = G(f_{ab1}),$$

where $b=a$ or $a+1$. By operating on both sides of (62) with an operator, which is the sum of all operators Ω_x and Ω_y for all sets of variables, it follows that, for all values of i and j , the form $f_{ijt} = F(f_{abs})$, where $b=a$ or $a+1$, $s=1, 2, \dots, e_b$. Hence, the quadratic forms w_{ijt} of (55) and (56) are functions of the forms

$$(63) \quad f_{ijt}, \quad (i=1, 2, \dots, k; j=i \text{ or } i+1, t=1, 2, \dots, e_j),$$

where f_{ijt} is defined by (59). There are $2n - e_1$ forms (63) and they are functionally independent, as they may be so arranged that each involves a variable which does not appear in any of its predecessors. Once again, e_1 is one half the degree of the minimal equation of HG^{-1} .

If H is non-singular, we can take H and G in diagonal block form $H = [H_1, H_2, \dots, H_t]$, $G = [G_1, G_2, \dots, G_t]$, where H_i and G_i are of order $2n_i$ and the elementary divisors of $H_i - xG_i$ are of one of the three types considered above. The number of functionally independent quadratic forms F is therefore $\sum_{i=1}^t (2n_i - 2m_i/2)$, where $2m_i$ is the degree of the minimal equation of $H_iG_i^{-1}$. But $\sum_{i=1}^t n_i = n$, and $\sum_{i=1}^t 2m_i = 2m$ is the degree of the minimal equation of HG^{-1} . Since the remaining case, in which H is

singular, is rather complicated, it is advisable to sum up our results in the form of

THEOREM 6. *Let H be the matrix of the Hamiltonian function of a linear conservative dynamical system with n degrees of freedom. If the degree of the minimal equation of HG^{-1} is $2m$, the system has $2n - m$ independent quadratic integrals unless H is singular.*

5. In order to obtain corresponding results for the case of a singular H , it is only necessary to consider the case in which every elementary divisor of $H - xG$ is of the form x^τ . We therefore consider the pencil $H - xG$, in which the elementary divisors are

$$x^{e_i}; \quad (i = 1, 2, \dots, k; e_1 \geq e_2 \geq \dots \geq e_k).$$

When τ is odd, the elementary divisor x^τ must occur an even number of times; hence, if e_i is odd, either e_i has the same value as e_{i-1} or the same value as e_{i+1} . From § 1 we see that a normal form for HG^{-1} is the diagonal block matrix

$$[W_1, W_2, \dots, W_k],$$

where $W_i = U_i$, if e_i is even, while, if e_i is odd, either $W_i = U_i$ and $W_{i+1} = -U'_i$, or $W_{i-1} = U_i$ and $W_i = -U'_i$. The normal form for G is also a diagonal block matrix with a block X_i (defined by (15)) corresponding to each even e_i and a block $\begin{pmatrix} 0 & E_i \\ -E_i & 0 \end{pmatrix}$ corresponding to each pair of odd e_i . In the above the matrices, E_i , U_i and X_i are the matrices of order e_i defined in § 1.

In order to determine the form of the most general symmetric matrix F , which satisfies (8), we let

$$(64) \quad C = FG^{-1}.$$

Since C is commutative with HG^{-1} , if $C = (C_{ij})$, $i, j = 1, 2, \dots, k$, we have

$$(65) \quad W_i C_{ij} = C_{ij} W_j.$$

The matrices C_{ij} are of four types depending on the structure of W_i and W_j . However, as we shall only be interested in symmetric matrices F , we need only consider the different possibilities for C_{ij} when $i \leq j$, so that $e_i \geq e_j$. Therefore, in what follows, i is always less than or equal to j . These possibilities are:

Type (i). $W_i = U_i$, $W_j = U_j$. Then $C_{ij} = \begin{pmatrix} G_{ij} \\ 0 \end{pmatrix}$, where G_{ij} is a polynomial in U_j .

Type (ii). $W_i = -U'_i$, $W_j = -U'_j$. Then $C_{ij} = \begin{pmatrix} 0 \\ K_{ij} \end{pmatrix}$, where K_{ij} is a polynomial in U'_j .

Type (iii). $W_i = U_i$, $W_j = -U'_j$. Since $U'_j X_j = -X_j U_j$, (65) becomes

$$U_i C_{ij} X_j = C_{ij} X_j U_j$$

and

$$C_{ij} = \begin{pmatrix} G_{ij} \\ 0 \end{pmatrix} X_j,$$

where $\begin{pmatrix} G_{ij} \\ 0 \end{pmatrix}$ is of *type (i)*.

Type (iv). $W_i = -U'_i$, $W_j = U_j$. Since $U_j X_j = -X_j U'_j$, (65) becomes

$$U'_i C_{ij} X_j = C_{ij} X_j U'_j,$$

so that

$$C_{ij} = \begin{pmatrix} 0 \\ K_{ij} \end{pmatrix} X_j, \text{ where } \begin{pmatrix} 0 \\ K_{ij} \end{pmatrix} \text{ is of type (ii).}$$

Symbolically we may denote a matrix C_{ij} of *type (p)* by T_p , $p = 1, 2, 3, 4$ and therefore symbolically have the result

$$(66) \quad T_3 = T_1 X \quad \text{and} \quad T_4 = T_2 X.$$

If $F = (F_{ij})$ is a partition of the matrix F , defined by (64), similar to that of C , the matrix F_{ij} is one of four distinct types.

If $W_i = U_i$ and e_i is even, $F_{ii} = C_{ii} X_i$. Since C_{ii} is of *type (i)*, by (66), F_{ii} is of *type (iii)*.

If $W_i = U_i$ and e_i is odd, $F_{ii} = -C_{i, i+1}$. Since e_i is odd, $W_{i+1} = -U'_i$ and $C_{i, i+1}$ and, therefore, F_{ii} is of *type (iii)*.

If $W_i = -U'_i$, $F_{ii} = C_{ii-1}$. Since $W_{i-1} = U_i$, the matrix C_{ii-1} and, therefore F_{ii} , is of *type (iv)*.

We accordingly have the lemma,

LEMMA 3. *The matrix F_{ii} is either of type (iii) or of type (iv). It is of type (iv) if, and only if, $W_i = -U'_i$. If F_{ii} is of type (iv), the matrices $F_{i-1, i-1}$ and $F_{i+1, i+1}$ are both of type (iii).*

e_j even: Then $F_{ij} = C_{ij} X_j$. If $W_i = U_i$, C_{ij} is of *type (i)* and, by (66), F_{ij} is of *type (iii)*. If $W_i = -U'_i$, C_{ij} is of *type (iv)* and, by (66), F_{ij} is of *type (ii)*. On replacing the matrices of $\begin{pmatrix} F_{ii} & F_{ij} \\ & F_{jj} \end{pmatrix}$ by their corresponding types, we may express the above results conveniently by the two diagrams

$$(67) \quad \begin{pmatrix} T_3 & T_3 \\ & T_3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} T_4 & T_2 \\ & T_3 \end{pmatrix}.$$

e_j odd, $W_j = U_j$: Then $F_{ij} = -C_{i,j+1}$. If $W_i = U_i$, F_{ij} is of type (iii) and, if $W_i = -U'_i$, F_{ij} is of type (ii). These two results lead to the same diagrams (67).

e_j odd, $W_j = -U'_j$: Then $F_{ij} = C_{i,j-1}$. If $W_i = U_i$, F_{ij} is of type (i) and, if $W_i = -U'_i$, F_{ij} is of type (iv). These results may be expressed by means of the two new diagrams

$$(68) \quad \begin{pmatrix} T_3 & T_1 \\ & T_4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} T_4 & T_4 \\ & T_4 \end{pmatrix}.$$

It is apparent from the diagrams (67) and (68) that the type of F_{ij} uniquely determines the types of F_{ii} and F_{jj} , and conversely. Since T_3 and T_4 involve the matrix X , while T_1 and T_4 do not, we shall call the matrices of types (i) and (ii) *positive*, and those of types (iii) and (iv) *negative*. For brevity we shall say that F_{ij} has sign ϵ , where $\epsilon = +1$ or -1 , according as F_{ij} is of positive or negative type.

$i+1$ less than j : If F_{ij} is positive, either F_{ii} is of type (iv) and F_{jj} of type (iii) or F_{jj} is of type (iv) and F_{ii} of type (iii). In the first of these cases, as a consequence of Lemma 3, $F_{i+1, i+1}$ is of type (iii), so that $F_{i, i+1}$ is of type (ii) and $F_{i+1, j}$ is of type (iii). In the second case, $F_{j-1, j-1}$ is of type (iii), $F_{i, j-1}$ of type (iii) and $F_{j-1, j}$ of type (i). We have therefore proved

LEMMA 4. If $i+1 < j$ and if F_{ij} is positive, there exists an integer k , $i < k < j$, such that F_{ik} and F_{kj} are of opposite sign.

On the other hand, if F_{ij} is negative, F_{ii} and F_{jj} are of the same type. Therefore, if $i < k < j$, F_{ik} and F_{kj} are both of the same sign. We may combine this last result with that of Lemma 4 to have

LEMMA 5. Let F_{ij} have the sign $\epsilon = \pm 1$ and let $i+1 < j$. Then there exists an integer k , $i < k < j$, such that, if the sign of F_{ik} is δ , the sign of F_{kj} is $-\epsilon\delta$.

We next obtain explicit formulae for the linearly independent quadratic forms. If F_{ii} is of type (iii), $F_{ii} = G_{ii}X_i$, where G_{ii} is a polynomial in U_i . Since F , and therefore F_{ii} , is symmetric, F_{ii} is an even or an odd polynomial in U_i , according as e_i is even or odd. Let

$$F_{iit} = U_i^{e_i-t} X_i.$$

Then, if $e_i = e$, and x is the vector with components x_1, x_2, \dots, x_e ,

$$x'F_{iit}x = g_{iit},$$

where

$$(69) \quad g_{iit} = \sum_{p=1}^t (-1)^p x_p x_{t+1-p}, \quad (t = 1, 2, \dots, e).$$

It is obvious from (69) that g_{iit} is zero, if t is even, so that there are only $[\frac{1}{2}(e+1)]$ linearly independent quadratic forms g_{iit} , for a fixed value of i . Similarly, if F_{ii} is of type (4), $F_{iit} = (U_i^{e-k})' X_i$ and $x' F_{iit} x = h_{iit}$, where

$$(70) \quad h_{iit} = \sum_{p=1}^t (-1)^{p+1} x_{e+1-p} x_{e+p-t}.$$

On comparing (69) and (70) we see that h_{iit} is of the same form as g_{iit} , if x_j is replaced by x_{e+1-j} .

Since F is symmetric, the linearly independent quadratic forms corresponding to the matrix F_{ij} are those whose matrices are of the form $\begin{pmatrix} 0 & F_{ij} \\ F'_{ij} & 0 \end{pmatrix}$. If $e_i = e$ and $e_j = d$ and x and y are vectors of dimensions e and d respectively, the corresponding quadratic form is

$$(71) \quad (x' y') \begin{pmatrix} 0 & F_{ij} \\ F'_{ij} & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2x' F_{ij} y.$$

Since, according to assumption, $e \geq d$, if F_{ij} is of type (iii), the linearly independent quadratic forms obtained from (71) are

$$(72) \quad g_{ijt} = \sum_{p=1}^t (-1)^{p+1} x_p y_{t+1-p}, \quad (t = 1, 2, \dots, d).$$

If F_{ij} is of type (iv) they are

$$(73) \quad h_{ijt} = \sum_{p=1}^t (-1)^{p+1} x_{e+1-p} y_{d+p-t}, \quad (t = 1, 2, \dots, d).$$

If F_{ij} is of type (i) they are

$$(74) \quad u_{ijt} = \sum_{p=1}^t x_p y_{d+p-t}, \quad (t = 1, 2, \dots, d),$$

and, if F_{ij} is of type (ii), they are

$$(75) \quad w_{ijt} = \sum_{p=1}^t x_{e+p-t} y_p, \quad (t = 1, 2, \dots, d).$$

Although at first sight it appears that there are four distinct types, there are really only two. If, in (73) and (74), we replace y_j by y_{d+1-j} and in (73) and (75), x_j by x_{e+1-j} , the forms h_{ijt} becomes the same as g_{ijt} while u_{ijt} and w_{ijt} both become

$$(76) \quad v_{ijt} = \sum_{p=1}^t x_p y_{t+1-p}, \quad (t = 1, 2, \dots, d).$$

The $2n$ original variables x may be placed in sets of e_1, e_2, \dots, e_k elements corresponding respectively to the symmetric matrices $F_{11}, F_{22}, \dots, F_{kk}$. If, when F_{ii} is of type (iv), the e_i variables associated with F_{ii} are relabelled in the reverse order, this relabelling is equivalent to the replacement of y_j by y_{d+1-j} and of x_j by x_{e+1-j} , the replacement by which (76) was obtained. All such replacements may be made simultaneously and, if this is done, the linearly independent quadratic forms corresponding to the matrices F_{ij} , $i \leq j$, $i, j = 1, 2, \dots, k$ are all of two types; the forms g_{ijt} of (72) and forms v_{ijt} of (76). It is of course apparent that, if $i = j$, the formula (72) reduces to (69). In conformity with our previous convention, the forms g_{ijt} are called negative, and the v_{ijt} positive. For convenience we shall denote these forms by f_{ijt} , so that

$$(77) \quad \text{either } f_{ijt} = v_{ijt} \text{ or } f_{ijt} = g_{ijt}.$$

Further, when there is no risk of confusion, we shall drop the suffixes i and j and write v_t , g_t and f_t for v_{ijt} , g_{ijt} and f_{ijt} respectively.

We now define two other linear differential operators, X and H , by

$$(78) \quad X = \sum_{i=1}^{c-1} (-1)^{i+1} i x_{i+1} \frac{\partial}{\partial x_i}, \quad H = \sum_{i=1}^{d-1} (-1)^{i+1} i y_{i+1} \frac{\partial}{\partial y_i}.$$

Then

$$\begin{aligned} (X + \epsilon H) v_t &= X \sum_{p=1}^t x_p y_{t+1-p} + \epsilon H \sum_{p=1}^t y_p x_{t+1-p} \\ &= \sum_{p=1}^t (-1)^{p+1} p x_{p+1} y_{t+1-p} + \epsilon \sum_{p=1}^t (-1)^{p+1} p y_{p+1} x_{t+1-p} \\ &= \sum_{p=1}^t (-1)^{p+1} p x_{p+1} y_{t+1-p} + \epsilon \sum_{p=0}^{t-1} (-1)^{t+1-p} (t-p) x_{p+1} y_{t+1-p} \\ &= \left. \begin{aligned} &= t \sum_{p=0}^t (-1)^{p+1} x_{p+1} y_{t+1-p} \\ &= t \sum_{p=1}^{t+1} (-1)^p x_p y_{t+2-p} = t g_{t+1} \end{aligned} \right\} \text{ if } \epsilon = (-1)^t. \end{aligned}$$

and therefore

$$(79) \quad (X + (-1)^t H) v_t = t g_{t+1}.$$

Further,

$$\begin{aligned} (X + \epsilon H) g_t &= X \sum_{p=1}^t (-1)^p x_p y_{t+1-p} + \epsilon (-1)^{t-1} H \sum_{p=1}^t (-1)^p y_p x_{t+1-p} \\ &= \sum_{p=1}^t (-1)^{2p+1} p x_{p+1} y_{t+1-p} + \epsilon (-1)^{t-1} \sum_{p=1}^t (-1)^{2p+1} p y_{p+1} x_{t+1-p} \\ &= - \sum_{p=1}^t p x_{p+1} y_{t+1-p} - \sum_{p=0}^t (t-p) x_{p+1} y_{t+1-p} \\ &= - \left. \begin{aligned} &= - t \sum_{p=0}^t x_{p+1} y_{t+1-p} = - t v_{t+1} \end{aligned} \right\} \text{ if } \epsilon = (-1)^{t-1}, \end{aligned}$$

and therefore

$$(80) \quad (X + (-1)^{t-1}H)g_t = -tv_{t+1}.$$

Since $x_1y_1 = v_1 = -g_1$,

$$\begin{aligned} (X - H)(x_1y_1) &= g_2, \\ (X - H)^2(x_1y_1) &= (X - H)g_2 = -2v_3, \\ (X - H)^3(x_1y_1) &= -2(X - H)v_3 = -3!g_4, \end{aligned}$$

and in general

$$(81) \quad \begin{aligned} (X - H)^{2s+1}(x_1y_1) &= (-1)^s(2s+1)!g_{2s+2}, \\ (X - H)^{2s}(x_1y_1) &= (-1)^s(2s)!v_{2s+1}. \end{aligned}$$

Similarly

$$(82) \quad \begin{aligned} (X + H)^{2s+1}(x_1y_1) &= (-1)^s(2s+1)!v_{2s+2}, \\ (X + H)^{2s}(x_1y_1) &= (-1)^{s-1}(2s)!g_{2s+1}. \end{aligned}$$

It was remarked earlier that the forms f_{ijt} of (77) are of two types, positive and negative. However, for a fixed i and j and variable t all f_{ijt} are of the same type, the type being determined by the sign of F_{ij} .

Since v_t is positive and g_t is negative, we may write (81) and (82) more compactly in the form

$$(83) \quad (X - \epsilon H)^{2s+1}(x_1y_1) = af_{2s+2}, \quad (X + \epsilon H)^{2s}(x_1y_1) = bf_{2s+1},$$

where $-\epsilon$ is the sign of f_t and a and b are numerical constants. If $i = j$, f_{iit} is the g_{iit} , which is defined by (69); and, on dropping the suffix i , we have, as the analogue of (83),

$$(84) \quad X^{2s}(x_1x_1) = ag_{2s+1} = af_{2s+1},$$

where a is a numerical constant.

We shall now prove that all the quadratic forms f_{ijt} are functions of those for which $j = i$ or $j = i + 1$. In so doing we shall say that f_{ijt} is *reducible*, if it is a function of quadratic forms f_{abc} , where either $b - a < j - i$ or $b = j$, $a = i$ and $c < t$. Clearly, $f_{ij1} = x_1y_1$, $i \neq j$, is reducible, since it is a function of $x_1^2 = f_{i11}$ and of $y_1^2 = f_{j11}$. The reducibility of f_{ij1} will be expressed by writing

$$f_{ij1} \equiv 0.$$

Thus, if $f \equiv h$, $f - h$ is *reducible*.

In showing that f_{ijt} is reducible, when $j > i + 1$, we shall use an induction proof. The proof consists of two essentially different parts, since the cases of even and odd values of t have to be treated separately. In fact, for odd t the restriction $j > i + 1$ may be replaced by the weaker inequality $j > i$.

We first assume that f_{ijt} , $j > i$ is reducible for $t \leq 2m$ and under this assumption prove that then $f_{ij\ 2m+1}$ is reducible. Since $(x_1 y_1)(x_1 y_1) = x_1^2 y_1^2$,

$$(X + \epsilon H)^{2m}(x_1 y_1)(x_1 y_1) = (X + \epsilon H)^{2m} x_1^2 y_1^2.$$

Therefore by Leibnitz' Theorem,

$$(85) \quad \sum_{r=0}^{2m} \binom{2m}{r} \{(X + \epsilon H)^r(x_1 y_1)\} (X + \epsilon H)^{2m-r}(x_1 y_1) \\ = \sum_{r=0}^{2m} \binom{2m}{r} \{X^r(x_1^2)\} H^{2m-r}(y_1^2).$$

If r is even, $X^r(x_1^2)$ and $H^{(r)}(y_1^2)$ are both reducible by (84). Moreover, if F_{ij} has the sign $-\epsilon$, and r is even but different from 0 or $2m$, $(X + \epsilon H)^r(x_1 y_1)$ and $(X + \epsilon H)^{2m-r}(x_1 y_1)$ are also reducible by (83) and our induction assumption. Accordingly, with this value of ϵ , (85) reduces to

$$(86) \quad 2(x_1 y_1)(X + \epsilon H)^{2m}(x_1 y_1) \equiv V - U,$$

where

$$(87) \quad U = \sum_{r \text{ odd}} \binom{2m}{r} \{(X + \epsilon H)^r(x_1 y_1)\} (X + \epsilon H)^{2m-r}(x_1 y_1)$$

and

$$V = \epsilon \sum_{r \text{ odd}} \binom{2m}{r} \{X^r(x_1^2)\} H^{2m-r}(y_1^2).$$

On writing

$$X^a(x_1) = X^a \quad \text{and} \quad H^b(y_1) = H^b,$$

we have in place of (87)

$$(88) \quad U = \sum_{r \text{ odd}} \binom{2m}{r} \sum_{a=0}^r \binom{r}{a} X^a (\epsilon H)^{r-a} \sum_{b=0}^{2m-r} \binom{2m-r}{b} X^b (\epsilon H)^{2m-r-b},$$

or

$$(89) \quad U = (2m)! \sum_{r \text{ odd}} \sum_{a=0}^r \sum_{b=0}^{2m-r} \frac{X^a X^b H^{r-a} H^{2m-r-b} \epsilon^{a+b}}{a! (r-a)! b! (2m-r-b)!}.$$

If U^* is obtained from U in (87) by replacing ϵ by $-\epsilon$, since r is odd, as a consequence of (83) and our induction assumption, U^* is reducible. Hence,

$$(90) \quad U \equiv U - U^*.$$

In (88) U is expressed in terms of powers of ϵ . In the difference $U - U^*$ all even powers will disappear and each odd power will occur with a factor two. Therefore, $U - U^*$ is equal to twice the sum of those terms on the right of (89) for which $a + b$ is odd. In this summation therefore each term has the factor $+\epsilon$. For fixed values of a and b , the coefficients of $2(2m)! \epsilon X^a X^b / a! b!$ is $\Sigma' H^{r-a} H^{2m-r-b} / (r-a)! (2m-r-b)!$, where the

summation extends over all odd values of r , for which $r \geq a$, $2m - r \geq b$. If a is odd, $r - a$ is even and, since $a + b$ is also odd, $2m - r - b$ is odd; while, if a is even, $r - a$ is odd and $2m - r - b$ is even. Hence in Σ' each term $H^p H^q / p! q!$, for which $p + q = 2m - a - b$, occurs exactly once, and therefore

$$\begin{aligned} \Sigma' H^{r-a} H^{2m-r-b} / (r-a)! (2m-r-b)! \\ = \frac{1}{2} \sum_{p+q=a+b} \binom{2m-a-b}{p} H^p H^q / (2m-a-b)! \\ = \frac{1}{2} H^{2m-a-b}(y_1 y_1) / (2m-a-b)! = \frac{1}{2} H^{2m-a-b}(y_1^2) / (2m-a-b)!. \end{aligned}$$

Therefore, on changing the order of summation in the sum for $U - U^*$, we have

$$\begin{aligned} U - U^* &= 2\epsilon(2m)! \sum_{a+b \text{ odd}} (X^a X^b / a! b!) \frac{1}{2} H^{2m-a-b}(y_1^2) / (2m-a-b)! \\ &= \epsilon \sum_{a+b \text{ odd}} \binom{2m}{a+b} X^{a+b}(x_1^2) H^{2m-a-b}(y_1^2) = V. \end{aligned}$$

Therefore, by (90), $U \equiv V$; while by (86), $(X + \epsilon H)^{2m}(x_1 y_1) \equiv 0$. But, by (83), $(X + \epsilon H)^{2m}(x_1 y_1) = a f_{ij \ 2m+1}$ and accordingly $f_{ij \ 2m+1}$ is reducible. Incidentally, we have shown that $f_{ij \ 2m+1}$ is a function of f_{iis} , f_{jjs} and f_{ijs} , where $s \leq 2m$.

We now show that, if $i + 1 < j$ and if f_{ijt} is reducible for $t \leq 2m + 1$, then $f_{ij \ 2m+2}$ is reducible. Let the sign of F_{ij} be ϵ and let z be the variable associated with the integer k of Lemma 4. Let δ be the sign of F_{ik} . Then, by Lemma 5, $-\delta\epsilon$ is the sign of F_{kj} . If Z is the differential operator obtained from X in (78) by replacing x by z , as a consequence of (83) and our induction assumption we have the following results:

$$\begin{aligned} (X + \epsilon H)^{2t+1}(x_1 y_1) &= a f_{ij \ 2t+2} \equiv 0, \text{ if } t < m, \\ (91) \quad (X - \epsilon H)^{2t}(x_1 y_1) &= b f_{ij \ 2t+1} \equiv 0, \text{ if } t < m, \\ (X + \delta Z)^{2t+1}(x_1 z_1) &\equiv (X - \delta Z)^{2t}(x_1 z_1) \\ &\equiv (\epsilon H - \delta Z)^{2t+1}(y_1 z_1) \equiv (\epsilon H + \delta Z)^{2t}(y_1 z_1) \equiv 0. \end{aligned}$$

Since $(x_1 y_1) z_1^2 = (x_1 z_1)(y_1 z_1)$,

$$(X + \epsilon H + \delta Z)^{2m+1}(x_1 y_1) z_1^2 = (X + \epsilon H + \delta Z)^{2m+1}(x_1 z_1)(y_1 z_1).$$

Therefore, by Leibnitz' Theorem, if $g = 2m + 1$,

$$\begin{aligned} \sum_{r=0}^g \binom{g}{r} \{ (X + \epsilon H)^r(x_1 y_1) \} (\delta Z)^{g-r}(z_1^2) \\ = \sum_{r=0}^g \binom{g}{r} \{ (X + \delta Z)^r(x_1 z_1) \} (\epsilon H + \delta Z)^{g-r}(y_1 z_1). \end{aligned}$$

On using the reduction relations (91), we have, as a consequence of this last equation,

$$z_1^2(X + \epsilon H)^g(x_1 y_1) \equiv V - U,$$

where

$$U = \sum_{r \text{ even}} \binom{g}{r} \{X + \epsilon H\}^r(x_1 y_1) \{\delta Z\}^{g-r}(z_1^2)$$

and

$$V = \sum_{r \text{ even}} \binom{g}{r} \{(X + \delta Z)^r(x_1 z_1)\}(\epsilon H + \delta Z)^{g-r}(y_1 z_1).$$

If V^* is obtained from V by replacing δ by $-\delta$, each term of V^* is reducible (see (91)). Therefore

$$V \equiv V - V^*.$$

On expanding V and V^* , we find

$$V - V^* = 2g! \sum_{r \text{ even}} \sum_{a=0}^r \sum_{b=0}^{g-r} \frac{X^a \epsilon^b H^b Z^{r-a} Z^{g-r-b} \delta^{g-a-b}}{a! b! (r-a)! (g-r-b)!},$$

where $g-a-b$ is odd, so that $a+b$ is even. On rearranging the order of summation, we find

$$\begin{aligned} V - V^* &= 2\delta \sum_{a+b \text{ even}} \binom{g}{a+b} \sum_{a=0}^{a+b} \binom{a+b}{b} X^a (\epsilon H)^b \frac{1}{2} Z^{g-a-b}(z_1^2) \\ &= \delta \sum_{r \text{ even}} \binom{g}{r} \{(X + \epsilon H)^r(x_1 y_1)\} Z^{g-r}(z_1^2) = U. \end{aligned}$$

Therefore $V - U \equiv 0$ and, accordingly, $f_{ij, 2m+2} \equiv 0$.

Since $f_{ij1} \equiv 0$, if $i \neq j$, it now follows that $f_{ijt} \equiv 0$ unless $j = i$ or $j = i + 1$; and that $f_{iit} \equiv 0$ when t is odd. Accordingly, all quadratic forms f_{ijt} are functions of the forms

$$(92) \quad f_{iit}, f_{i-1, is}; \quad t \text{ odd and } s \text{ even}; \quad t, s \leq e_i.$$

If $e_i = 2m_i$ or $2m_i - 1$, the number of forms f_{iit} in (92) is m_i . The number of forms $f_{i-1, is}$ in (92) is m_i , if $e_i = 2m_i$, and is $m_i - 1$, if $e_i = 2m_i - 1$. The total number of forms $f_{iit}, f_{i-1, is}$ in (92), for a fixed $i \neq 1$, is therefore e_i . Accordingly, the total number l of forms (92) is

$$l = m_1 + \sum_{i=2}^k e_i = m_1 + 2n - e_1,$$

where $e_1 = 2m_1$ or $e_1 = 2m_1 - 1$.

Hence

$$l = 2n - m_1, \text{ if } e_1 = 2m_1, \text{ and } l = 2n + 1 - m_1, \text{ if } e_1 = 2m_1 - 1.$$

That these l forms (92) are actually functionally independent is apparent, if they are written in the order

$$f_{111}, f_{113}, \dots; f_{122}, f_{124}, \dots; f_{221}, \dots; f_{kk1}, \dots,$$

since each of these forms contains at least one variable which does not appear in any of its predecessors. The minimal equation of HG^{-1} is of degree e_1 , and, therefore, the number l can be expressed in terms of e_1 and the order of HG^{-1} . By the argument used to prove Theorem 5 we can now complete the proof of the theorem,

THEOREM 7. *Let H be the matrix of the Hamiltonian function of a linear conservative dynamical system with n degrees of freedom. If e is the degree of the minimal equation of HG^{-1} , the system has $2n - \left\lfloor \frac{e}{2} \right\rfloor$ functionally independent quadratic integrals.*

6. Linear integrals. If $y'x = \sum_{i=1}^{2n} y_i x_i = l$ is a linear integral of (6), condition (2) reduces to

$$y'GHx \equiv 0 \text{ or } HGy = 0.$$

Conversely, if $HGy = 0$, l is a linear integral of (6). Since G is non-singular we have the result ²²—a linear integral of the system (6) exists if, and only if, H is singular.

If l is a linear integral, l^2 is a quadratic integral, for

$$(l_x^2)'GHx = 2y'gHx \equiv 0.$$

The number of linearly independent linear integrals is k , where $2n - k$ is the rank of HG or HG^{-1} or H . We may express this by

THEOREM 7 bis.²³ *The number of linearly independent linear integrals is the number k of the elementary divisors of the form x^r belonging to the pencil $H - xG$.*

With the notation of the previous section the linearly independent linear integrals are

$$\sqrt{f_{ii1}}; \quad (i = 1, 2, \dots, k).$$

There is one and only one integral for each value of i . If e_i is even,

$$\sqrt{f_{ii1}} = x_{\sigma+1}, \quad \sigma = e_1 + e_2 + \dots + e_{i-1};$$

²² This result is not new. See Aurel Wintner, "On the linear conservative dynamical systems," *Annali di Matematica pura ed applicata*, ser. 4, tomo 13 (1934-35), pp. 105-112.

²³ This theorem is proved by Wintner. See reference ²².

if e_j is odd but $W_j = U_j$ (so that $e_{j+1} = e_j$), then

$$\sqrt{f_{jj1}} = x_{\rho+1} \text{ and } \sqrt{f_{j+1,j+1,1}} = x_\tau, \quad \rho = e_1 + e_2 + \dots + e_j, \quad \tau = \rho + e_j$$

Since G is now in the normal form of § 1 and is therefore a diagonal block matrix, and since $x_{\sigma+1}$ is the only linear integral corresponding to its block in G , $x_{\sigma+1}$ is in involution with all other linear integrals. The same is true of $x_{\rho+1}$ and x_τ . Further, the integrals $x_{\rho+1}$ and x_τ are in involution unless $e_j = 1$. In this last case, $e_j = 1$, they are not in involution. Hence we have the theorem

THEOREM 5 bis. *The number of linearly independent integrals in involution consists of $k - f$ members, where k is the total number of elementary divisors of the pencil $H - xG$ which have the form x^r and $2f$ is the number of those for which $r = 1$.*

As a consequence of the above theorem we have the corollary,

COROLLARY 1. *All linear integrals of the system (6) are in involution if, and only if, the pencil $H - xG$ has no linear elementary divisor of the form $x - 0$.*

Further, if there exist f linearly independent pairs of linear integrals which are not in involution, the pencil $H - xG$ has f pairs of linear elementary divisors of the form $x - 0$, $x - 0$.

Obviously linear independence of linear forms is synonymous with functional independence.

7. In the particular case of the small vibrations about an equilateral Lagrangian libration point in the restricted problem of three bodies,²⁴ the matrix H of the Hamiltonian function is

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1/4 & -z \\ 1 & 0 & -z & -5/4 \end{pmatrix}, \text{ where } z = \frac{\sqrt{27}}{4} (1 - 2\mu).$$

This matrix can be written more conveniently in the form $\begin{pmatrix} e & i \\ -i & k \end{pmatrix}$, where

$$e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad k = \begin{pmatrix} 1/4 & -z \\ -z & -5/4 \end{pmatrix}.$$

²⁴ Gylden, *Bull. Astr.*, vol. 1 (1884), pp. 361-369; E. Strömberg, "Über die kritische Masse im problème restreint und über das problème restreint im allgemein," *Publikationer og mindre meddelelser fra Københavns Observatorium*, Nr. 72 (1930).

The characteristic equation of HG^{-1} is

$$(93) \quad x^4 + x^2 + \frac{2\gamma}{4}\mu(1-\mu) = 0,$$

and the roots of this equation are given by

$$(94) \quad x^2 = \frac{-1 \pm \sqrt{1 - 2\gamma\mu + 2\gamma\mu^2}}{2}.$$

If the radical in (94) is negative the four roots of (93) are all complex and distinct while, if the radical is positive, the four roots, while still distinct, are all purely imaginary. When the radical is zero the roots of (94) are all imaginary but are equal in pairs. In this, the critical case, the general solution contains secular terms.²⁵ Nevertheless, as far as quadratic integrals are concerned, this case is the same as the other two. For, in the critical case, the characteristic equation of HG^{-1} is $(x^2 + \frac{1}{2})^2 = 0$. Since

$$(HG^{-1})^2 = \begin{pmatrix} -e-k & -2i \\ -i & -e-k \end{pmatrix},$$

the minimal equation of HG^{-1} is not $x^2 + \frac{1}{2} = 0$. Accordingly, for all values of μ , the minimal equation of HG^{-1} is the same as its characteristic equation. Therefore, by Theorem 1, there are exactly two independent quadratic integrals: the energy integral whose matrix is H and the integral whose matrix is $(HG^{-1})^2H$. A simple calculation shows that

$$(HG^{-1})^2H = \begin{pmatrix} -3e-k & -ik \\ ki & -e - \frac{2\gamma}{4}\mu(\mu-1)e \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{13}{4} & \frac{\sqrt{2\gamma}}{4}(1-2\mu) & \frac{\sqrt{2\gamma}}{4}(1-2\mu) & \frac{5}{4} \\ \frac{\sqrt{2\gamma}}{4}(1-2\mu) & -\frac{\gamma}{4} & \frac{1}{4} & -\frac{\sqrt{2\gamma}}{4}(1-2\mu) \\ \frac{\sqrt{2\gamma}}{4}(1-2\mu) & \frac{1}{4} & -\frac{2\gamma}{4}\mu(\mu-1)-1 & 0 \\ \frac{5}{4} & -\frac{\sqrt{2\gamma}}{4}(1-2\mu) & 0 & -\frac{2\gamma}{4}\mu(\mu-1)-1 \end{pmatrix}.$$

THE JOHNS HOPKINS UNIVERSITY.

²⁵ A. Wintner, "Librationtheorie des restringierten Dreikörperproblems," *Mathematische Zeitschrift*, vol. 32 (1930), pp. 660-661.

ERRATA.

In the joint paper, J. E. Eaton and Oystein Ore: *Remarks on multi-groups*, in the January number of this JOURNAL, pp. 67-71, the following condition has inadvertently been omitted in the definition of *proper homomorphism*:

3. If $m_1^* \cdot m_2^*, m_3^*$ then for some m_1, m_2, m_3 one has $m_1 m_2, m_3$.

Mr. R. S. Pote of the University of Illinois has called our attention to this fact.

In the paper, J. E. Eaton, *Associative multiplicative systems*, in the January number of this JOURNAL, pp. 222-232, the following statement appearing in § 5, p. 225 is incorrect:

The X_i 's then obviously form an m -system which is homomorphic to the original m -system.

Dr. H. H. Campaigne of the University of Minnesota has called attention to this error.

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

ABRAHAM COHEN
THE JOHNS HOPKINS UNIVERSITY

F. D. MURNAGHAN
THE JOHNS HOPKINS UNIVERSITY

T. H. HILDEBRANDT
UNIVERSITY OF MICHIGAN

J. F. RITT
COLUMBIA UNIVERSITY

R. L. WILDER
UNIVERSITY OF MICHIGAN

WITH THE COÖPERATION OF

OYSTEN ORE
H. P. ROBERTSON
M. H. STONE
T. Y. THOMAS
G. T. WHYBURN

E. T. BELL
H. B. CURRY
E. J. MCSHANE
HANS RADEMACHER
OSCAR ZARISKI

C. R. ADAMS
R. D. JAMES
SAUNDERS MACLANE
GABOR SZEGÖ
LEO ZIPPIN

PUBLISHED UNDER THE JOINT AUSPICES OF
THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXII, Number 4
OCTOBER, 1940

THE JOHNS HOPKINS PRESS
BALTIMORE, MARYLAND
U. S. A.

CONTENTS

	PAGE
On the non-existence of the Euclidean algorithm in certain quadratic number fields. By ALFRED BRAUER,	697
Postulational bases for the umbral calculus. By E. T. BELL,	717
The abelian quasi-group. By HARRIET GRIFFIN,	725
The Gaussian law of errors in the theory of additive number theoretic functions. By P. ERDŐS and M. KAC,	738
On the standard deviations of additive arithmetical functions. By PHILIP HARTMAN and AUREL WINTNER,	743
On the almost periodicity of additive number-theoretical functions. By PHILIP HARTMAN and AUREL WINTNER,	753
On the spherical approach to the normal distribution law. By PHILIP HARTMAN and AUREL WINTNER,	759
On upper limit relations for number theoretical functions. By PHILIP HARTMAN and RICHARD KERSHNER,	780
On the properties of a collective. By Z. W. BIRNBAUM and HERBERT S. ZUCKERMAN,	787
On symmetric Bernoulli convolutions. By TATSUO KAWATA,	792
The four-vertex theorem for spherical curves. By S. B. JACKSON,	795
A complete characterization of sectional families of curves. By ANNETTE VASSELL,	813
Exactly $(k, 1)$ transformations on connected linear graphs. By O. G. HARROLD, JR.,	823
The characterization of pseudo-spherical sets. By LEONARD M. BLUMENTHAL and GEORGE R. THURMAN,	835
A geometry associated with Cremona's equations. By GERALD B. HUFF,	855
Polynomials whose real part is bounded on a given curve in the complex plane. By A. C. SCHAEFFER and G. SZEGÖ,	868
Neuer beweis eines Satzes von G. H. Hardy und S. Ramanujan über das asymptotische Verhalten der Zerfallungskoeffizienten. Von VOJISLAV G. AVAKUMOVIĆ,	877
An algebraic problem involving the involutory integrals of linear dynamical systems. By JOHN WILLIAMSON,	881
Errata,	912

THE AMERICAN JOURNAL OF MATHEMATICS will appear four times yearly.

The subscription price of the JOURNAL for the current volume is \$7.50 (foreign postage 50 cents); single numbers \$2.00.

A few complete sets of the JOURNAL remain on sale.

Papers intended for publication in the JOURNAL may be sent to any of the Editors.

Editorial communications may be sent to Professor F. D. MURNAGHAN at The Johns Hopkins University.

Subscriptions to the JOURNAL and all business communications should be sent to THE JOHNS HOPKINS PRESS, BALTIMORE, MARYLAND, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

THE JOHNS HOPKINS PRESS • BALTIMORE

- American Journal of Mathematics.** Edited by ABRAHAM COHEN, T. H. HILDEBRANDT, F. D. MURNAGHAN, J. F. RITT and R. L. WILDER. Quarterly. 8vo. Volume LXII in progress. \$7.50 per volume. (Foreign postage, fifty cents.)
- American Journal of Philology.** Edited by H. CHERNISS, K. MALONE, B. D. MERITT, and D. M. ROBINSON. Quarterly. 8vo. Volume LXI in progress. \$5 per volume. (Foreign postage twenty-five cents.)
- Biologia Generalis. (International Journal of Biology).** Founded by LEOPOLD LÖHNER, Graz; RAYMOND PEARL, Baltimore, and VLADISLAV RŮŽIČKA, Prague. It is now edited by O. ABEL, L. ADAMETZ, O. PORSCH, C. SCHWARZ, J. VERSLUYS and R. WASICKY of Vienna, 8vo. Volume XV in progress.
- Bulletin of the History of Medicine.** Edited by HENRY E. SIGERIST. Monthly except August and September. Volume VIII in progress. 8vo. Subscription \$5 per year. (Foreign postage, fifty cents.)
- Bulletin of the Johns Hopkins Hospital.** Edited by JAMES BORDLEY, III. Monthly. Volume LXVII in progress. 8vo. Subscription \$6 per year. (Foreign postage, fifty cents.)
- Comparative Psychology Monographs.** ROY M. DORCUS, Managing Editor. 8vo. Volume XVI in progress. \$5 per volume.
- Hesperia.** Edited by WILLIAM KURRELMEYER and KEMP MALONE. 8vo. Thirty-two numbers have appeared.
- Human Biology: a record of research.** RAYMOND PEARL, Editor. Quarterly. 8vo. Volume XII in progress. \$5 per volume. (Foreign postage, thirty-five cents.)
- Johns Hopkins Studies in Romance Literatures and Languages.** H. C. LANCASTER, Editor. 8vo. Forty-nine numbers have been published.
- Johns Hopkins University Circular,** including the President's Report and Catalogue of the School of Medicine. Ten times yearly. 8vo. \$1 per year.
- Johns Hopkins University Studies in Archaeology.** DAVID M. ROBINSON, Editor. 8vo. Twenty-nine volumes have appeared.
- Johns Hopkins University Studies in Education.** FLORENCE E. BAMBERGER, Editor. 8vo. Twenty-eight numbers have appeared.
- Johns Hopkins University Studies in Geology.** EDWARD B. MATHEWS, Editor. 8vo. Thirteen numbers have been published.
- Johns Hopkins University Studies in Historical and Political Science.** Under the direction of the Departments of History, Political Economy and Political Science. 8vo. Volume LVIII in progress. \$5 per volume.
- Modern Language Notes.** Edited by H. C. LANCASTER, W. KURRELMEYER, R. D. HAVENS, K. MALONE, H. SPENCER and C. S. SINGLETON. Eight times yearly. 8vo. Volume LV in progress. \$5 per volume. (Foreign postage, fifty cents.)
- Reprint of Economic Tracts.** J. H. HOLLANDER, Editor. Fifth series in progress. Price \$4.
- Terrestrial Magnetism and Atmospheric Electricity.** Founded by LOUIS A. BAUER; conducted by J. A. FLEMING with the cooperation of eminent investigators. Quarterly. 8vo. Volume XLV in progress. \$3.50 per volume.
- Walter Hines Page School of International Relations.** Eight volumes have been published.

A complete list of publications will be sent upon request

THE THEORY OF GROUP REPRESENTATIONS

By Francis D. Murnaghan

We have attempted to give a quite elementary and self-contained account of the theory of group representations with special reference to those groups (particularly the symmetric group and the rotation group) which have turned out to be of fundamental significance for quantum mechanics (especially nuclear physics). We have devoted particular attention to the theory of group integration (as developed by Schur and Weyl); to the theory of two-valued or spin representations; to the representations of the symmetric group and the analysis of their direct products; to the crystallographic groups; and to the Lorentz group and the concept of semi-vectors (as developed by Einstein and Mayer).

—Extract from Preface.

380 pages, 8vo, cloth, \$5.00

NUMERICAL MATHEMATICAL ANALYSIS

By JAMES B. SCARBOROUGH

"A valuable feature of the book is the excellent collection of examples at the end of each chapter. . . . The book has many admirable features. The explanations and derivations of formulae are given in detail. . . . The author has avoided introducing new and complicated notations which, although they may conduce to brevity, are a serious stumbling block to the reader. The typography and paper are excellent."

—*American Mathematics Monthly*.

430 pages, 25 figures, crown 8vo, buckram, \$5.50

TABLES OF $\sqrt{1-r^2}$ AND $1-r^2$ FOR USE IN PARTIAL CORRELATION AND IN TRIGONOMETRY

By JOHN RICE MINER, Sc. D.

These tables fill a want long felt by practical workers in all branches of statistics. Everyone who uses the method of correlation has wished for tables from which the probable error of a coefficient of correlation could be obtained *with accuracy*. Similar tables to this have existed on a small scale, but never before have there been available tables of $1-r^2$ and $\sqrt{1-r^2}$ to 6 places of decimals, and 4 places in the argument. Not only are these tables of great usefulness in getting the probable error of a correlation coefficient, but also they have what will perhaps be their chief value in the calculations involved in the method of partial or net correlation. It is safe to say that these tables reduce the labor involved in this widely used statistical method by at least one-half.

50 pages, 8vo, cloth, \$1.50

THE JOHNS HOPKINS PRESS · BALTIMORE

